The conservation and divergence of function in the SIX family of homeodomain transcription factors

Raphaëla Joanna Kitson-Pantano

A thesis presented for the degree of PhD, University of Edinburgh, 2007

Declaration

This thesis has been composed by myself and the work presented herein is my own, except where stated.

Raphaëla Joanna Kitson-Pantano

I dedicate this thesis to Joanie Oliver Kitson.

Acknowledgements

I thank the Jarman, Finnegan and Armstrong labs and particularly my supervisors Prof. Andrew Jarman and Prof. David Finnegan for welcoming me and providing me with the facilities and academic support I needed to succeed in this PhD. I also thank the Darwin Trust for funding my PhD and Principal Tim O'Shea for his support.

I owe a big thank you to all the people who never doubted me and were there for me throughout this PhD. Thank you for your patience, kindness and confidence in me. This includes, to name but a few, (in alphabetical order) **Jonathan Bard**, **Jamie Davies**, **Darren Downing**, **Philippe Josse, Anthi Kiose, Tom Mackenzie**, **Carol Noble**, **The Reynolds family, Anne Rush, Stuart Sayer**, **Jo Young**, and all the people in Pollock, staff and Fresher since 2004, particularly **Lawrence Head** who was so supportive during the last few weeks of the writing up.

Thank you to (in alphabetical order) **Aileen, Charlotte, Derek, Diane, Fozia, Frances, Isabelle**, **Jeff, John, Laure, Maria, Pete, Ramani, Vernon,** my doctors **Dr. Ganne** and **Dr. Usdin** and my loving Godmother **Nikky**. I could never have managed without you. You all mean the world to me. You are the pillars of my strength, the best friends I have, the family I chose. Thank you so much.

Donald, you know the real me. You supported me in all my projects. I would never have succeeded without your friendship. Thank you for keeping my secrets and respecting my choices.

Theo, you were the happiness in this PhD. These three and a half years were ours. We did this together. I will never forget it. You are the best friend I have ever had. Thank you.

My Dearest Mummy, this PhD is yours as much as mine. "We did it!" Je n'y serais jamais arrivée sans toi. Thank you for putting up with my bad moods, my tantrums and just generally with me. Thank you for always encouraging me, supporting me and for always believing in me. Thank you for being the best mother in the world and for loving me always. I love you best in the whole wide world. Na na na ni na!

♣♣♣♣♣

Abstract

The *Six* genes encode a group of transcription factors that are characterised by a Six domain believed to be required for protein interaction, and a homeodomain necessary for DNA binding. Bioinformatics analysis, carried out in this PhD, shows the evolutionary sequence conservation of these genes from poriferans to mammals. There are three *Six* genes in *Drosophila* and six *Six* genes in mammals. *Six* genes in mammals are involved in myogenesis and neurogenesis. SIX1, SIX4 and SIX5 are coexpressed in mouse myogenesis and SIX1 and SIX4 are expressed in sensory neurons early in trigeminal gangliogenesis. SIX3 is expressed in the developing eye and the forebrain and ectopic expression of SIX3 results in the formation of the retina. *Six* genes are also involved in human diseases. SIX1 is involved in the branchio-oto-renal syndrome, SIX3 in holoproencephaly and the disruption in expression of the *SIX5* gene is thought to be involved in myotonic dystrophy.

The *Drosophila melanogaster* homologue of *SIX5*, *D-Six4,* when mutated results in embryonic defects both in muscle and gonad development which are thought to mirror the symptoms of myotonic dystrophy patients which include the inability to relax muscles after contraction, cataracts, mental deficiencies and sterility. *D-Six4* is part of a family of three *Six* genes (*D-Six4, sine oculis, Optix*). Mutations in these genes result in different phenotypes. Sine oculis and Optix are expressed in the head and the developing eye. They are required for *Drosophila* eye development. D-Six4 is expressed in the mesoderm, it is required for muscle and gonad developments. Sequence identity, observed in the Six domain and homeodomain of *Six* gene orthologues, has led to the grouping of Six proteins into three subfamilies, Six1/2, Six4/5 and Six3/6.

The aims of this PhD were to determine whether *Drosophila* is a good model for studying *Six* genes and to investigate the conservation and divergence of function of the different *Six* genes in *Drosophila melanogaster.* Firstly, I carried out sequence comparisons of different Six proteins from poriferans to mammals and identified features of conservation with putative functional importance. Importantly, I proposed specific criteria defining the Six domain more accurately. I then hypothesised that the different SIX proteins are functionally distinct and that these functional differences are conserved between species. In order to test this, the ability of one Six protein to substitute for the loss of function of another Six protein was assessed through genetic rescue experiments. I found that Optix and Sine oculis can substitute for the loss of function of D-Six4 in the muscle but not very efficiently and Sine oculis can partially compensate for loss of D-Six4 in the gonad. This also gave insight into the role of *D-Six4* in mesoderm development. Finally, I sought to investigate the ability for a vertebrate orthologue to substitute for a mutation of a *Drosophila Six* gene but due to experimental difficulties that could not have been anticipated this was not achieved. While the ability of only the three *Six* genes of *Drosophila* to complement the muscle and gonad phenotypes in a *D-Six4* mutant was assessed, molecular and fly work carried out using mammal *Six1*, *Six4, Six5* and *Six3* DNA resulted in invaluable learning outcomes with regards to my training as a scientist.

The first three results chapters, chapter 2, 3 and 4, are bioinformatics analyses of the protein sequence comparisons carried out throughout the metazoan phylum. The ability for the *Drosophila melanogaster* paralogues of D-Six4 (Sine oculis and Optix) to rescue aspects of the *D-Six4* mutant phenotype are then discussed in

chapter 5. Chapter 6 discusses the work that was carried out in attempting to generate molecular constructs with the vertebrate orthologues eg. *Six1*, *Six4* and *Six5* genes.

♣♣♣♣♣

Table of contents

Chapter 1

Introduction

Introduction

1.1 Why study Developmental Biology?

Developmental Biologists are interested in the development of living organisms throughout their life cycle. Using molecular biology, genetics and cell biology technologies we research the intricacies that lead to different cell shapes, patterns and behaviours. Our recent ability to mutate and introduce new genes into embryos and to suppress gene expression in specific cells at specific times in development has revolutionised the study of Developmental Biology. "One of the most exciting aspects of Developmental Biology is that understanding a developmental process in one organism can help illuminate similar processes elsewhere" (Wolpert, 1998). *Drosophila melanogaster* (*D. melanogaster*) is the paradigm example of this as indeed our understanding of its development and the genetic basis of it has led to the discovery of related and functionally similar genes in mammals and other vertebrates (Wolpert, 1998). It is from such discoveries that general developmental principles emerge (Wolpert, 1998).

Personally, I chose to study Developmental Biology because of its beauty. Developmental Biology combines the challenging molecular biology techniques and the interesting complexity of genetics with the fascinating work with living organisms and the beautiful images obtained through confocal microscopy.

1.2. Why study the Six family of transcription factors?

The first *Six* gene to be discovered was *sine oculis* (*so*) in *D. melanogaster* (Cheyette et al., 1994). As I will discuss further on, So is a key protein involved in *D. melanogaster* eye development (Cheyette et al., 1994). Its paralogues in *D.*

melanogaster and orthologues in mammals (from here onwards meaning mouse and humans unless otherwise specified) play critical roles in several aspects of development as I explain further. The *Six* genes are therefore important genes involved in development and consequently interesting to study.

The *Six* genes encode proteins that function as transcription factors and are characterised by a distinctive homeodomain (HD) (60 amino acids (aa)) that has diverged considerably from previously described HDs (Cheyette et al., 1994). The HD is known to confer DNA binding specificity (Cheyette et al., 1994). "The Six proteins HD is unique in that it lacks two highly conserved aa residues typical of most HDs, an arginine at position 5 and a glutamine at position 12" (Cheyette et al., 1994). These proteins also contain a Six domain (SD) (116 aa) at the amino-terminus of the HD which is thought to facilitate protein-protein interactions (Cheyette et al., 1994).

1.3 Why use D. melanogaster as a model?

D. melanogaster is an ideal model organism. Its lifecycle is short (10-12 days), it is cheap and easy to keep, its genome has been fully sequenced (Adams et al., 2000), mutagenesis and genetic studies are easy to carry out. "Of the 289 genetic flaws known to cause disease in humans, *Drosophila* homologues for 60% have been found" (Rubin et al., 2000) An understanding of the *Six* genes in *D. melanogaster* is essential for understanding *Six* genes in other organisms.

1.4 What is known of Six genes in D. melanogaster?

D. melanogaster has three *Six* genes, *so, Optix* and *D-Six4. so* and *Optix* are most well-known for their requirement in eye development whereas *D-Six4* is principally involved in muscle and gonad development. As I will discuss later, this thesis focuses on *D-Six4* and its role in muscle and gonad development. Below, I briefly discuss the role of *so* and *Optix* in eye development so as to emphasise the functional differences with regards to *D-Six4.*

1.4.1 What is the function of so in eye development?

so mutants analysis shows that it is required for the development of the entire visual system, which includes the adult compound eye, the ocelli, the optic lobe and the larval photoreceptor organ (Cheyette et al., 1994). In the eye disc, So is expressed before the development of the morphogenetic furrow (MF) and is thought to have an essential role in the initiation of pattern formation as many *so* mutant discs display a failure of any ommatidia to develop (Cheyette et al., 1994). So is also thought to be required for propagation of the furrow since rescue of the loss-of-function mutations necessitate the continuous supply of the *so* gene product (Cheyette et al., 1994). The role of the *so* gene in the eye disc is thought to be similar to another eyeless mutation, *eya* (Cheyette et al., 1994). Eya is also expressed anterior to the MF and molecular epistasis suggests that *eya* acts prior to *so* (Cheyette et al., 1994).

In the developing optic lobes, "So is first expressed early on, just prior to invagination, in the region of the embryonic ectoderm that will give rise to the optic primordium (OP)" (Cheyette et al., 1994). In *so* mutants, the OP fails to invaginate and extensive cell death occurs in the optic lobes during the first half of pupal

development (Cheyette et al., 1994). So is a member of a group of transcription factors that together direct the formation of the eye (Silver and Rebay, 2005). This group is known as the retinal determination gene network (RDGN) (Silver and Rebay, 2005). "Within this context, So and Eya are thought to act as a complex to regulate common steps in eye development" (Pignoni et al., 1997). They interact in yeast and *in vitro* the N-terminal domain of Eya interacts with the Six domain of So (Pignoni et al., 1997).

For over a decade ectopic expression of So alone was thought to have little or no effect on antennal, wing, or leg disc development (Pignoni et al., 1997; Weasner et al., 2007). In contrast to previous reports, Weasner *et al* recently showed that the expression of So, on its own, is sufficient to initiate eye development in non-retinal tissues (the antennal disc) (Weasner et al., 2007). They showed that the induction of ectopic eyes does not require the co-expression of Eya as previously held (Weasner et al., 2007). These recent findings indicate an ability for So to function independently of Eya and puts it on equal footing with the second *D. melanogaster Six* gene *Optix* that is also able to function independently of Eya.

1.4.2 How was Optix discovered and how does it function in comparison to So in eye development?

In search of orthologues of the mammal Six proteins, *Drosophila* cDNA clones were isolated and sequence comparisons between homeodomains led to the identification of *Optix* (Toy et al., 1998). *Optix* maps to a position relatively close to the *so* locus suggesting that these two *Six* genes arose by a tandem gene duplication event (Seimiya and Gehring, 2000; Toy et al., 1998). To date, aside from *so*, examination of the *Fly Base* registry has not yet revealed the existence of any *Drosophila* eye

mutants that map near the *Optix* locus. Optix is thought to have important functions in eye development since ectopic expression of Optix leads to the formation of ectopic eyes (Seimiya and Gehring, 2000).

Optix and *so* belong to the same gene family and their protein sequences share a high degree of amino acid sequence identity (discussed later in more detail). The expression patterns of Optix and So are different. In the developing eye imaginal disc; *Optix* expression is restricted to cells ahead of the morphogenetic furrow while *so* is expressed throughout the eye field (Seimiya and Gehring, 2000). So functions synergistically with Eya in eye development whereas Optix does not (Seimiya and Gehring, 2000). These results suggest that the developmental roles of So and Optix are different (Seimiya and Gehring, 2000).

Recently it was shown that within the antennal disc, So and Optix can both induce ectopic eyes albeit in different locations of the antenna (Weasner et al., 2007). Outside of the antenna, Optix is expressed in the wing and haltere imaginal discs and ectopic eyes can be generated by Optix (Weasner et al., 2007). In contrast, the ability of *so* to induce ectopic eyes appears restricted to the antennal disc and its adult derivatives (Weasner et al., 2007).

1.4.3 How do the domains of So and Optix specify their different functions?

Recent work has been carried out demonstrating the different domain capabilities of these two Six proteins So and Optix (Weasner et al., 2007). Specificity in partner selection was shown to be crucial to the role that So and Optix play in eye development (Weasner et al., 2007). Indeed, rescue of the mutant eye phenotype by a So protein with a deleted or replaced (by an Optix SD) SD was not observed (Weasner et al., 2007). These results suggest that the SD regions of So and Optix

bind to different partners (Weasner et al., 2007). Yeast two hybrid assays using the SD domains of the Six proteins as baits have identified differing sets of putative binding partners for So and Optix (Weasner et al., 2007). The transcriptional co-repressor Groucho binds both So and Optix (Kenyon et al., 2005) whereas Eya and SBP (So binding protein) interact strongly with So but not with Optix (Eya binds very weakly to Optix)(Kenyon et al., 2005). Inversely, Obp (Optix binding protein) binds preferentially to Optix (Kenyon et al., 2005).

Weasner *et al* focused their study of the functional conservation of these domains solely on So and Optix (Weasner et al., 2007). They published these results toward the end of my PhD. As the title of this thesis reads, the focus of my PhD has been the conservation and divergence of function of the Six genes. My work therefore includes functional comparisons between So and Optix but also extends the comparisons to the third *D. melanogaster* gene, *D-Six4*, and *Six* genes in other species. I will now discuss how D-Six4 is involved in the patterning of the mesoderm.

1.4.4 How was D-Six4 discovered and what is its role in the patterning of the mesoderm?

D-Six4 was discovered by using degenerate primers derived from the C-termini of the Six domains and homeodomains of several Six class proteins and carrying out PCR experiments to amplify related sequences on *Drosophila* larval cDNA (Seo et al., 1999). Its homeodomain was thus identified and subsequently, the full-length of the gene was isolated by applying a RACE strategy (Seo et al., 1999). Following this discovery, *in situ* hybridisation experiments were carried out in order to decipher the

expression pattern and potential function of D-Six4 (Seo et al., 1999). However, it was only following an EMS screen and the subsequent isolation of two mutations that mapped to the *D-Six4* chromosomal location, failed to complement each other, and produced embryos that exhibited a mutant phenotype consistent with the D-Six4 expression pattern that the function of D-Six4 was more clearly understood (Kirby et al., 2001).

1.4.4.1 What is the expression pattern of D-Six4?

During embryogenesis, *D-Six4* mRNA is expressed in the developing head region, the CNS and the mesoderm (Seo et al., 1999). At stage 9; expression in the mesoderm is generally transient and coincides with Mef2, a protein that is expressed throughout the mesoderm at this stage (Clark et al., 2006; Kirby et al., 2001). By stage 10, *D-Six4* mRNA becomes restricted to ventral and lateral mesoderm, being lost in the dorsal mesoderm, before being rapidly lost from all mesodermal cells except the somatic gonadal precursors (SGPs) (Clark et al., 2006; Kirby et al., 2001). At stage 13 mesodermal expression becomes segmental and confined weakly to the SGPs in parasegments 10–12, which subsequently form the somatic sheath that surrounds the gonad (Clark et al., 2006; Kirby et al., 2001). At stage 15, *D-Six4* expression then becomes strong in the SGPs after they have coalesced with the migrating germ cell precursors (pole cells) to form the immature gonad (Clark et al., 2006; Kirby et al., 2001).

1.4.4.2 What did mutation analysis reveal?

As I mentioned above, an EMS screen resulted in the isolation of two mutations that mapped to the *D-Six4* chromosomal location (Kirby et al., 2001). The first mutation, *D-Six4*^{289} a nonsense point mutation (C_{1753} > T), gives rise to a stop codon in place of Gln_{87} (Kirby et al., 2001), in the N-terminal domain. It is a likely null mutation (Kirby et al., 2001). The second mutation, *D-Six4*^{131} a point mutation (C_{2404} > T), results in an amino acid substitution of Cys for Arg_{281}, and is less severe (Kirby et al., 2001). This mutation will be further discussed in chapter 2 as it maps to a critical part of the Six domain. *D-Six4*^{131} mutant embryos hatch normally, although many die during larval and pupal stages (Kirby et al., 2001). A small proportion survives to adulthood (Kirby et al., 2001). These mutations have contributed to understanding how D-Six4 functions. *D-Six4* appears to be a key factor for the development of a variety of tissues that originate from the non dorsal mesoderm (Clark et al., 2006). It is required for fat body precursors, SGPs and the lateral and ventral muscles (Clark et al., 2006; Kirby et al., 2001). It is likely to be a competence factor or patterning mediator, acting to regulate a variety of key tissue and cell identity genes (Clark et al., 2006; Kirby et al., 2001). Below I explain how muscle and gonads develop during *D. melanogaster* embryogenesis and I further discuss the role of D-Six4 in the development of these processes.

1.4.4.3 How do muscles develop in *D. melanogaster* embryogenesis?

Most muscles derive from the development of the mesoderm which is a complex and dynamic process. The patterning of the mesoderm involves the interaction of many transcription factors, the key players being *twist, tinman* and *Mef2*. In an attempt to keep this introductory chapter as precise and as concise as possible, I shall first discuss the involvement of these three crucial transcription factors in the patterning of the mesoderm and will then explain how individual muscles are derived.

1.4.4.3.1 How do twist, tinman and Mef2 pattern the mesoderm?

The *Drosophila twist* gene, which encodes a bHLH transcription factor, is initially expressed within the presumptive mesoderm, where it patterns the dorsoventral (D–V) axis with the transcription factor (TF) Dorsal (stage 5) (Thisse et al., 1987), (Shirokawa and Courey, 1997). Dorsal activates *twist*, which in turn coregulates the majority of known direct Dorsal targets (Sandmann et al., 2007). Afterward, it drives the process of gastrulation together with the TF Snail (stage 6) (Ip et al., 1992). Within the unspecified mesoderm, Twist acts as a master regulator that is both essential and sufficient to initiate mesoderm development (stages 7–11) (Thisse et al., 1987). Once the mesoderm primordium is specified and differentiation begins, Twist expression is dramatically reduced and is only maintained in adult muscle precursors (Bate et al., 1991).

Many genes have been identified that are genetically downstream from *twist* and almost 500 direct target genes of Twist have been identified by microarray analysis (Sandmann et al., 2007). These include *tinman*, and *Mef2*. Twist regulates an almost complete cassette of genes required for morphogenesis in addition to genes

essential for cell communication, signal transduction, cell motility, cell adhesion, cell proliferation and cell migration (Sandmann et al., 2007). Twist is also thought to have a role in the establishment or maintenance of the anterior-posterior patterning within the mesoderm as some of its gene targets include the gap, pair rule, and segmentation genes, as well as homeotic genes (Sandmann et al., 2007). Twist potentially contributes to the regulation of the majority of TFs involved in every aspect of early mesoderm development and is essential for the invagination of mesoderm precursors during gastrulation, segmentation, and specification of muscle types (Sandmann et al., 2007) .

twist is required for the expression of the homeobox gene *tinman* which is expressed in the mesoderm primordium (Azpiazu and Frasch, 1993). Later in development, as the first mesodermal subdivisions are occurring, expression of Tinman becomes limited to the visceral mesoderm and the heart progenitors (Bodmer, 1993). The function of *tinman* is required for visceral muscle and heart development (Bodmer, 1993). Embryos that are mutant for the *tinman* gene lack the appearance of visceral mesoderm and of heart primordia, and the fusion of the anterior and posterior endoderm is impaired (Bodmer, 1993). Even though *tinman* mutant embryos do not have a heart or visceral muscles, many of the somatic body wall muscles appear to develop although abnormally (Bodmer, 1993). When the *tinman* cDNA is ubiquitously expressed in *tinman* mutant embryos, via a heatshock promoter, formation of heart cells and visceral mesoderm is partially restored, *tinman* seems to be one of the earliest genes required for heart development and the first gene reported for which a crucial function in the early mesodermal subdivisions has been implicated (Bodmer, 1993). *tinman* is not expressed in *twist* mutants but *twist* is

not the sole activator of *tinman, decapentaplegic (dpp)* is required for the maintenance and enhancement of *tinman* in tissues that become the precursors of the heart (Frasch, 1995). In addition to its dependence on *dpp*, dorsal mesodermal *tinman* expression requires the activity of *tinman* itself, as *tinman* mutant embryos show strongly reduced expression (Frasch, 1995).

Mef2 encodes a *Drosophila* homolog of the vertebrate myocyte-specific enhancer factor 2 (MEF2) (Bour et al., 1995). It is expressed throughout the mesoderm following gastrulation (Bour et al., 1995). Later in embryogenesis, its expression is maintained in precursors and differentiated cells of the somatic and visceral musculature, as well as the heart (Bour et al., 1995). Complete loss of *Mef2* protein in homozygous mutant embryos exhibit a dramatic absence of myosin heavy chain (MHC)-expressing myoblasts and lack differentiated muscle fibers (Bour et al., 1995). Examination of earlier events of muscle development indicates that the specification and early differentiation of somatic muscle precursors are not affected (Bour et al., 1995). However, these partially differentiated cells are unable to undergo further differentiation to form muscle fibers in the absence of *Mef2*. The later aspects of differentiation of the visceral mesoderm and the heart are also disrupted in *Mef2* mutant embryos, although the specification and early development of these tissues appear unaffected (Bour et al., 1995). Midgut morphogenesis is disrupted in the mutant embryos, presumably as a consequence of abnormal development of the visceral mesoderm. In the heart, the cardial cells do not express MHC (Bour et al., 1995). These results indicate that *Mef2* is required for later aspects of differentiation of the three major types of musculature, which include body wall muscles, gut musculature, and the heart, in the *Drosophila* embryo (Bour et al.,

1995). A temporal map of Mef2 activity during *Drosophila* embryonic development has more recently been devised and the number and diversity of new direct target genes indicate a broad role for Mef2, at all stages of myogenesis (Sandmann et al., 2006). Cooperation with *twist* has been suggested (Sandmann et al., 2006).

1.4.4.3.2 How do muscles develop?

Twist appears to target both *tinman* and *Mef2* which are involved in the development of the muscles. *Mef2* in particular is important in later stages of muscle differentiation. Muscles originate from the fusion of two types of myoblasts, founder cells (FCs) and fusion-competent myoblasts (FCMs) (Rushton et al., 1995). Each of the 30 body wall muscles in an abdominal hemisegment is initiated by a single FC and each acquires unique identities controlled by its own unique combination of transcriptional factors, including *Kruppel, S59, vestigial;* and *apterous,* necessary for the development of a particular muscle (Bate and Rushton, 1993; Bourgouin et al., 1992; Dohrmann et al., 1990; Ruiz-Gomez, 1998; Ruiz-Gomez et al., 1997). By contrast to FCs, FCMs are thought to be a naïve group of myoblasts, entrained to a particular muscle program upon fusion to the FC. These two groups of myoblasts have distinct transcriptional profiles and FCMs possibly have a greater role in determining the final morphology of the muscle (Ruiz Gomez and Bate, 1997). Genes required for myoblast fusion include *dumbfounded* which is involved in the recognition event between founder cells and fusion-competent myoblast and *myoblast city* which is involved in the event of fusion itself (Ruiz-Gomez et al., 2000; Rushton et al., 1995).

1.4.4.3.3 What effect does D-Six4^{289} mutation have the development of muscles?

Wild-type embryos when stained at stage 16 with an antibody to Myosin display a regular pattern of myotubes (Kirby et al., 2001). In homozygous *D-Six4*^{289} mutant embryos stained with an antibody to Myosin, the cardioblasts and the dorsal somatic muscles are normal (Clark et al., 2006). Lateral and ventral somatic muscles however are strongly disrupted and some muscles are missing, including the lateral transverse muscles (LT) (Clark et al., 2006). The number and location of missing muscles varies between segments and between embryos and identifying all the remaining muscles in embryos stained for Myosin is very difficult. Nonetheless a number of consistent effects can be discerned (Clark et al., 2006). In the ventral region, certain muscles are consistently affected to a greater extent than others (Clark et al., 2006). The ventral acute muscle, VA3, and segment border muscle (SBM) are absent in most segments, while some ventral oblique (VO) muscles (VO1-3) and most of the ventral longitudinal (VL) muscles are usually present (Clark et al., 2006). Many isolated rounded myosin-expressing cells are scattered among the muscles (Clark et al., 2006). These presumably are myoblasts that have not fused with elongated founder cells attempting to form a myotube (Clark et al., 2006). "The requirement of *D-Six4* for the development of specific muscles that arise from the dorsolateral and ventral regions suggests that *D-Six4* regulates specific muscle identity genes in founder myoblasts (Clark et al., 2006). Consistent with this proposal, the expression of *ladybird*, which is a muscle identity gene for the SBM founder myoblast, is absent in *D-Six4* mutant embryos" (Clark et al., 2006; Jagla et al., 1997; Ruiz-Gomez, 1998). The expression of dorsal muscle identity genes such as *collier* and *eve* is

unaffected however (Clark et al., 2006). These results support an active role for *D-Six4* in initial patterning of cell fates (Clark et al., 2006). It is possible, therefore, that maintenance/survival phenotypes are a secondary effect of defects in the initial establishment of cell identity (Clark et al., 2006).

1.4.4.3.4 How does D-Six4 interact with the other transcription factors involved in muscle development?

In at least some aspects of somatic muscle patterning, the roles of *tinman* and *D-Six4* are complementary (Clark et al., 2006). However, the relationship between *D-Six4* and *tinman* changes over time and *tinman* has functions in the ventral and lateral mesoderm that remain obscure (Clark et al., 2006). In the dorsal mesoderm at stage 10/11, *D-Six4* expression is complementary to that of *tinman*, and there are no discernable effects on dorsal mesoderm structures in *D-Six4* mutants (Clark et al., 2006; Lee and Frasch, 2005). It is thought that these two genes play complementary roles promoting the development of specific cell types in conjunction with additional patterning factors (Clark et al., 2006). *tinman* mutant embryos show specific lateral and ventral muscle defects that presumably depend on its early pan-mesodermal expression and are different from that of *D-Six4* mutants (Azpiazu and Frasch, 1993). Muscles affected by *tinman* include LL1, LO1, VL3, VL4, and VT1 which do not require D-Six4 (Azpiazu and Frasch, 1993). Conversely, muscles that are severely affected by *D-Six4* mutation appear normal in *tinman* mutants, including VA3, the SBM, and the external lateral muscles LT1, LT2, LT3, and LT4 (Azpiazu and Frasch, 1993).

Similarly to *D-Six4*, *eya* expression is first restricted to the dorsolateral and ventral mesoderm and later to the SGPs (Boyle et al., 1997; Clark et al., 2006). Unlike *D-Six4*, however, *eya* expression is not modulated in the anteroposterior axis (Clark et al., 2006). Mesodermal phenotypes of *eya* and *D-Six4* mutants are similar for SGP and muscle development (Boyle et al., 1997; Clark et al., 2006; Moore et al., 1998). Misexpression of *eya* alone has little or no effect on mesodermal organs (Boyle et al., 1997; Clark et al., 2006) however comisexpression of *eya* with *D-Six4* in the *twist* domain strongly affects the mesoderm (Clark et al., 2006). Regular muscle pattern is formed but several cell types of the ventral and dorsolateral mesoderm are expanded suggesting that *eya* with *D-Six4* have a patterning role (Clark et al., 2006). In the ventral region, there are consistently duplications of the VA3 muscle, while the pattern of VO muscles is disrupted (Clark et al., 2006). This suggests that at least some of the muscles that are disrupted in the *D-Six4* mutant are duplicated upon misexpression, whereas muscles that are little affected in the mutant (dorsal and some ventral) are disrupted by misexpression (Clark et al., 2006). The lateral somatic musculature is also expanded upon misexpression (Clark et al., 2006). "These results support the idea of distinct developmental pathways for muscle subsets, one of which requires *D-Six4* (Clark et al., 2006). Significantly, other pathways (some requiring *tinman*) are susceptible to *D-Six4/eya* function" (Clark et al., 2006).

1.4.4.3.5 What is left to be discovered?

I have given an overview of the main factors involved in the development of somatic muscles in *D.melanogaster*. *twist* and its direct downstream targets *tinman* and *Mef2* play a critical role in this process. D-Six4 is involved in this complex network of

interacting transcription factors. Many questions arise about not only how these transcription factors interact, but which ones do, when and for what purpose. Understanding how D-Six4 functions will shed light on some of these questions.

1.4.4.4 How do gonads develop?

D-Six4 also plays a major role in gonad formation (Kirby et al., 2001). Gonads are formed from three distinct cell types: primordial germ cells (PGCs), somatic gonadal precursors (SGPs), and in males, male-specific somatic gonadal precursors (msSGPs) (Boyle and DiNardo, 1995; DeFalco et al., 2003; Santos and Lehmann, 2004). These originate in distinct locations and migrate to associate in two intermingled clusters which then compact to form the spherical primitive gonads (Boyle and DiNardo, 1995; DeFalco et al., 2003; Santos and Lehmann, 2004).

The *Drosophila* PGCs are specified at the posterior pole of the embryo and are carried into the gut cavity during gastrulation, before migrating actively through the midgut epithelium (Santos and Lehmann, 2004). The cells diverge bilaterally away from the midline as they migrate along the basal surface of the midgut and then detach to move to the lateral mesoderm (Sano et al., 2005). At around the same time, the somatic gonadal precursors (SGPs) are specified in the dorsolateral mesoderm on either side of the embryo in parasegments 10, 11 and 12 (Boyle and DiNardo, 1995). During germ band retraction, the three groups of SGPs coalesce with each other by mid stage 12 to form a single cluster intermingled with germ cells by the start of stage 13 (Clark et al., 2007). The compaction of the gonads occurs symmetrically, with SGPs moving both anteriorly and posteriorly towards a central focus (Clark et al., 2007). Recently, it was discovered that another group of somatic cells are

important. The msSGPs are specified in both sexes in parasegment 13 in a position ventral to the SGPs (DeFalco et al., 2004). They migrate anteriorly and dorsally as a single tight cluster to join the posterior of the gonad in males; while in females they are eliminated by apoptosis during this migration (DeFalco et al., 2004). D-Six4 is expressed in the SGPs from their first appearance until at least the end of embryogenesis and is consequently a candidate regulator of SGP function during gonad formation (Clark et al., 2007).

1.4.4.4.1. What effects does the D-Six4 mutation have?

Embryos homozygous for the hypomorphic mutation D-*Six4*^{131} show a severe defect in gonadogenesis (Clark et al., 2007; Kirby et al., 2001). SGPs are present and appear to be correctly specified but the three parasegmental SGP clusters fail to merge to form a unified gonad, indicating a failure in SGP migration or mutual recognition (Clark et al., 2007; Kirby et al., 2001). Also PGCs do not associate with them and become scattered (Kirby et al., 2001). Similarly, the migration of msSGPs to the gonad is impaired (Clark et al., 2007; Kirby et al., 2001). Most mutant SGPs retain the ability to associate with each other and with PGCs over short distances, but there is a defect in their ability to coalesce over longer distances (Clark et al., 2007; Kirby et al., 2001). A hypothesis is that in D-*Six4*^{131}, signalling is disrupted, and therefore compaction may occur via stochastic D-Six4-independent local cell contacts (Clark et al., 2007; Kirby et al., 2001). This would account for the variability in whether the anterior or posterior SGP cluster fails to be incorporated into the gonad-like structure (Clark et al., 2007; Kirby et al., 2001). An alternative explanation for the SGP coalescence phenotype of D-*Six4*^{131} is that fully functional D-Six4 is required in

SGPs for the compaction process to operate consistently over the distance between SGP clusters, perhaps by regulating the length of productive cellular protrusions that may be required for SGP-SGP contacts (Clark et al., 2007; Kirby et al., 2001). SGPs in embryos lacking fully functional D-Six4 fail to coalesce to form unified gonads (Kirby et al., 2001). D-Six4 is thus a key regulator of somatic cell function during gonadogenesis (Kirby et al., 2001). Many of the SGPs that initially appear in *D-Six4*^{289} mutants appear to die subsequently (Clark et al., 2007; Kirby et al., 2001). Thus, *D-Six4* appears to be required in SGPs both for initial specification and for survival or maintenance of fate (Clark et al., 2007; Kirby et al., 2001). In *D-Six4*^{289} homozygous embryos, initial germ cell internalization and migration are normal, but the cells then fail to coalesce to form a gonad (Kirby et al., 2001). D-Six4 is also required for msSGP migration (Clark et al., 2007). The msSGPs must migrate a substantial distance to reach the developing gonad (Clark et al., 2007). Because D-Six4 is expressed strongly in msSGPs, as well as SGPs, it has not yet been determined if the failure of msSGPs to migrate reflects a defect in the msSGPs themselves or defective signalling by the SGPs (Clark et al., 2007).

1.4.4.4.2 What are the genes regulating gonadogenesis?

The specification of SGPs requires the functions of the homeotic gene *abdominal A* (*abd-A*) and *eya* (Boyle and DiNardo, 1995). *eya*, whose expression is refined through negative regulation by *bagpipe*, acts together with *abd-A* to specify anterior gonadal precursor fates in parasegment 10 and 11 (Boyle and DiNardo, 1995). In *abd-A* mutants, Eya fails to be expressed, and in embryos overexpressing Abd-A, Eya is expressed ectopically (Boyle and DiNardo, 1995). Despite the early

expression of Eya, SGPs are specified in the absence of Eya function but they fail to maintain their fate (Boyle and DiNardo, 1995). Consequently, germ cells do not coalesce into a gonad and thus mutations in *eya* abolish gonad formation and produce female sterility (Boyle and DiNardo, 1995). Abd-A also acts together with Abd-B to specify a posterior subpopulation of gonadal precursors (Boyle and DiNardo, 1995). 9-12 cells are selected as SGPs within each of three posterior parasegments at early stages in gonadogenesis (Boyle and DiNardo, 1995; Clark et al., 2007). In addition, initial dorsoventral positioning of SGPs relies on a regulatory cascade that establishes dorsal fates within the mesoderm (Clark et al., 2006). *tinman* appears to mediate the role of ectodermally expressed *decapentaplegic* (Clark et al., 2006; Frasch, 1995; Staehling-Hampton et al., 1994). In *tinman* mutants few or no SGP cells are detected. The anteroposterior and dorsoventral position of the SGP cells within a parasegment also rely on the ectodermly secreted growth factor wingless (wg) (Clark et al., 2006; Lee and Frasch, 2005). While loss of *wg* abolishes SGPs, ectopic expression expands the population such that most cells within lateral mesoderm adopt gonadal precursor fates (Clark et al., 2006; Lee and Frasch, 2005). Mutation analysis shows that *D-Six4* is required in SGPs initial specification as SGPs in *D-Six4*^{289} mutants express the markers *412*, *eya* and *zfh-1* but it remains unclear how it interacts with the other transcription factors (Clark et al., 2007). In wild type embryos, all areas of *D-Six4* expression coincide with an area of Eya expression (Kirby et al., 2001). At stage 13; they are coexpressed in SGPs suggesting that D-Six4 is a partner of Eya in these tissues (Kirby et al., 2001). In *D-Six4*^{289} mutants the number of cells expressing Eya decreases markedly during stage 12, while a greater number of apoptotic cells is observed in the region normally occupied by SGPs

(Kirby et al., 2001). Some of these apoptotic cells also express Eya, implying that programmed cell death accounts for the reduction in SGP number at this time (Kirby et al., 2001).

Following specification, initial association of SGP clusters can occur where SGPs fail to coalesce into a single structure suggesting that the two processes differ mechanistically under distinct control from the movements that drive gonad compaction, perhaps dependent on distinct short- or long-range interactions respectively, the former being D-Six4-independent (Clark et al., 2007). D- Six4 is required for expression of *Hmgcr* which codes for *HMGCoA* reductase and is necessary for attraction of PGCs by SGPs (Clark et al., 2007). However SGP signalling is unlikely to be *Hmgcr*-dependent, since mutation of *Hmgcr* does not prevent clusters of SGPs from associating, even though they do not attract germ cells (Clark et al., 2006). In *Hmgcr* mutants it is clear that the movements of gonad compaction follow the initial association of the three parasegmental SGP clusters, implying that these two processes represent distinct cell behaviours at different developmental stages (Clark et al., 2007). It seems likely therefore that *Hmgcr* is one of the target genes that must be regulated by D-Six4 for normal SGP function. Regulation of *Hmgcr* is unlikely to be the only function of D-Six4 in the SGPs, since gonad formation takes place in *Hmgcr* mutant embryos (Clark et al., 2006; Clark et al., 2007). Communication between SGP clusters during coalescence possibly requires a D-Six4-dependent signalling mechanism operating between SGPs (Clark et al., 2007). This may direct the migration of the three clusters towards each other prior to compaction, or it may influence the polarity of the compaction process, ensuring that a single gonadal cluster is formed (Clark et al., 2007). D-Six4 would

positively regulate a component of this signalling pathway, either in the signalling or in the receiving SGPs, or both (Clark et al., 2007). The mechanism attracting msSGPs to the gonads is distinct from that attracting the PGCs as they associate correctly with SGPs in *Hmgcr* mutant male embryos (Clark et al., 2007). Also expressed in the msSGPs is *dsxm* although its role remains unclear (Clark et al., 2007). Following gonad formation, male germline divisions, regulated by the *Jak/STAT* pathway, begin, whereas there are no divisions of female germ cells at this stage (Clark et al., 2007; Wawersik et al., 2005). A few mutations affect late stages of gonad morphogenesis and all of these are believed to affect the adhesive properties of SGPs (Clark et al., 2007).

1.4.5 How does D-Six4 function in comparison to So and Optix?

D-Six4 is expressed in different tissues to So and Optix and drives the development of very different organs. Nonetheless all three genes are members of the same family of transcription factors with significant amino acid similarity. This raises questions about what makes these proteins different. If So and Optix were expressed in the mesoderm would they be capable of carrying out similar functions? Eya is a cofactor of So in the development of the eye, and is also coexpressed with D-Six4 in the mesoderm. This suggests that Eya and D-Six4 might interact. Are the So-Eya interactions similar or even interchangeable with D-Six4-Eya interactions? Optix can function in an Eya-independent fashion. Does this make it a better candidate than So for carrying out D-Six4 functions in the mesoderm?

Other transcription factors are involved both in the development of the muscle and gonad. Are So and Optix able to interact with these transcription factors

in similar ways? Would these transcription factors have an inhibitory effect on these proteins or vice versa? Is what makes proteins different determined by their expression patterns, their cofactors, their binding sites, a combination of these and other factors?

There has been an apparent separation of function between the So and Optix proteins and D-Six4 in *D. melanogaster*. It is relevant to question whether this distinction in function is conserved in other species. How ancient are the functions carried out by the different SIX subfamilies and when did the subfamilies arise?

1.5 What is known of Six genes in other species?

Mammals have six *Six* genes (*Six1-Six6*) which can be classed into the same three subfamilies as the *Drosophila* genes (Seo et al., 1999). There are two orthologues for every one in *D. melanogaster* (Seo et al., 1999). Phylogenetic analysis suggests that *Six1/Six2*, the orthologues of *so*, are more closely related to *Six4/Six5*, the orthologues of *D-Six4* than to *Six3/Six6* (Seo et al., 1999). An understanding of these proteins expression patterns and function in mammals might help decipher whether the observed segregation of function between the *Six* genes in *D. melanogaster* is relatively recent or represents a primordial condition in the common ancestor.

Similarly to *Drosophila, Six* genes are not only involved in muscle and gonad development in mammals. Again in an attempt to remain both precise and concise, I will briefly discuss what is known about *Six* genes in humans and in mouse in general. I will then focus on the role of *Six* genes in myogenesis so as to emphasise similarities and differences with *Drosophila.*

1.5.1 What is known about Six genes in humans?

Human embryonic SIX3 is expressed in the eye field and SIX6 is expressed in the developing and adult human retina (Seo et al., 1999). SIX3 is also thought to have wide implications in head development (Seo et al., 1999). In contrast, human SIX1, SIX4 and SIX5 homeoproteins are expressed in myogenic cells and together with EYA proteins control the early activation of myogenic regulating factors such as MYF5, myogenin and MRF4 (Fougerousse et al., 2002). These factors are potentially involved in later steps of myogenic differentiation (Fougerousse et al., 2002). The involvement of human SIX4/SIX5 in myogenin expressing cells is comparable to *Drosophila* D-Six4 involvement in the development of the muscle and may suggest a conservation of function of the Six4/Six5 subfamily in the development of muscle structures.

In humans, mutations in *SIX* genes are the cause of disease. An example of this is, Myotonic Dystrophy (DM1) an autosomal-dominant multisystem disease characterized by progressive skeletal muscle wasting and weakness, myotonia, cataracts, cardiac arrhythmias, mild mental retardation, insulin resistance, and endocrinopathies including testicular and tubular atrophy, oligospermia, Leydig cell hyperproliferation and increased follicle stimulating hormone (FSH) levels (Johnson et al., 1996). DM1 is the most common inherited neuromuscular disease in adults (Johnson et al., 1996). In DM1 patients, transcriptional silencing of the SIX5 allele on chromosome 19 through the expansion of a flanking and unstable CTG repeat located in the 3' untranslated region of the myotonic dystrophy protein kinase gene (Dmpk), which encodes a serine-threonine protein kinase, is observed (Sato et al., 2002). There are three mechanisms by which CTG expansion can result in DM1

(Inukai et al., 2000). First, repeat expansion may alter the processing or transport of the mutant DMPK mRNA and consequently reduce DMPK levels (Inukai et al., 2000). Second, CTG expansion may establish a region of heterochromatin 3' of the repeat sequence and decrease SIX5 transcription (Inukai et al., 2000). Third, toxic effects of the repeat expansion may be intrinsic to the repeated elements at the level of DNA or RNA (Inukai et al., 2000). Dysfunction of SIX5 is thought to be primarily responsible for the ophthalmic features of DM1

The role of SIX5 in the development of these tissues is very different to that of D-Six4 in *Drosophila* where D-Six4, as discussed, is principally involved in the patterning of the mesoderm as opposed to SIX5 being involved in eye development. However, myotonic dystrophy patients display muscle defects and testicular atrophy and it has been suggested that while the exact phenotypic relationship between D-Six4 and SIX4/5 remains to be elucidated, the defects in *D-Six4* mutant flies suggest that human SIX5 should be more strongly considered as being responsible for the muscle wasting and testicular atrophy phenotypes in DM1 (Kirby et al., 2001).

Other diseases caused by mutations in the *Six* genes include the branchio-oto-renal (BOR) syndrome, an autosomal-dominant disorder characterized by hearing loss, branchial and renal anomalies (Kochhar et al., 2007) and holoprosencephaly (HPE), a severe brain malformation which results from incomplete cleavage of the forebrain during early embryogenesis (Wallis et al., 1999). BOR syndrome is the result of mutations in *SIX1* and *SIX5*, together with or independently of mutations in *EYA1* (Kochhar et al., 2007) and in some patients with HPE, *SIX3* has been found to be mutated in the homeodomain (Wallis et al., 1999). The involvement of SIX3 in

the development of the eye is comparable to that of Optix in *Drosophila* suggesting a possible conservation of function in this subfamily.

Overall there may be a conservation of function of the Six4/Six5 subfamily in the development of muscle structures as well as a conservation of function of the Six1/Six2 subfamily and Six3/Six6 subfamily in the development of the eye both in invertebrates and vertebrates. Additionally, the Six4/Six5 subfamily seems to carry out functions in the development of the eye in vertebrates. The interactions SIX1-EYA1 and So-Eya may also be conserved.

1.5.2 What is known about Six genes in mouse?

Six homeoproteins are expressed in several mouse tissues and are involved in diverse differentiation processes including neurogenesis and the development of other organs.

Six1-/- neonates have strongly disorganised craniofacial structures including the inner ear, the nasal cavity, the craniofacial skeleton, and the lacrimal and parotid glands (Ozaki et al., 2004). *Six1-/-* mice also lack a thymus and a parathyroid (Ozaki et al., 2004). Analyses of the thymus/parathyroid development in *Six1-/-;Six4-/-* double mutants show that Six1 and Six4 act synergistically to control morphogenetic movements of early thymus/parathyroid tissues (Ozaki et al., 2004) indicating a collaboration in function between Six1 and Six4 that is not observed in *Drosophila.* Six1 together with Six2 play an essential role in early kidney development (Xu et al., 2003) (Brodbeck et al., 2004). While *Six4*-deficient mice have no apparent abnormalities, *Six1/Six4*-deficient mice exhibit more severe kidney phenotypes than

the *Six1*-deficient mice (Xu et al., 2003). Again, Six4 and Six1 are involved together in the development of trigeminal ganglia (Konishi et al., 2006).

Six3 plays a crucial role in early eye and forebrain development (Oliver et al., 1995). *Six3* mRNA is found in structures including the ectoderm of nasal cavity, olfactory placode and Rathke's pouch, and also the ventral forebrain including the region of the optic recess, hypothalamus and optic vesicles (Oliver et al., 1995). Contrarily to the other members of the *Six* gene family, Six3 does not interact with any known member of the Eya family and does not function as a transcriptional activator rather Six3 acts as a potent transcriptional repressor upon its interaction with Groucho-related members (Zhu et al., 2002). Six6 may also be involved in vertebrate eye development (Zhu et al., 2002) but distinct expression patterns support the idea that Six 3 and Six 6 are differentially required during forebrain development (Jean et al., 1999).

1.5.3 What is the involvement of these genes in myogenesis?

Six5, *Six4* and *Six1* genes are co-expressed during mouse myogenesis (Grifone et al., 2005). *Six1-/-* mice die at birth due to thoracic skeletal defects and severe muscle hypoplasia affecting most of the body muscles (Laclef et al., 2003). *Six1-/-* embryos have impaired primary myogenesis, characterized, at E13.5, by a severe reduction and disorganisation of primary myofibers in most body muscles (Laclef et al., 2003). *Six1* defects are very similar to phenotypes caused by mutations of *Eya1* (Laclef et al., 2003). Close comparison of *Six1* and *Eya1* deficient mice strongly suggests a functional link between these two factors (Laclef et al., 2003). *Pax* gene mutations also lead to comparable phenotypes, suggesting that a regulatory network including

the *Pax, Six* and *Eya* genes is required for several types of organogenesis in mammals (Ikeda et al., 2002). Six1 has been shown to synergise with Eya2 to regulate myogenesis (Fougerousse et al., 2002). Direct physical interactions have shown synergistic regulation of muscle development by Dach2 with Eya2 and Eya2 with Six1 and parallel the synergistic regulation of *Drosophila* eye formation by dachshund with Eya and Eya with So (Fougerousse et al., 2002; Ikeda et al., 2002). This genetic network is used in a novel developmental context, myogenesis rather than eye development, and has been expanded to include gene family members that are not directly homologous, for example Pax3 instead of Pax6 (Fougerousse et al., 2002; Ikeda et al., 2002). *Six1* role in muscle is not conserved in *Drosophila.*

Six4 single knockout (KO) mice have no developmental defects (Ozaki et al., 2001). *Six1Six4* double KO mice however show an aggravation of the phenotype previously described for the single *Six1* KO (Grifone et al., 2007; Grifone et al., 2005). At the limb bud level, *Six1* and *Six4* homeogenes control early steps of myogenic cell delamination and migration from the somite through the control of Pax3 gene expression(Grifone et al., 2007). Absence of *Six1* and *Six4* impairs the expression of the myogenic regulatory factors myogenin and Myod1, and Mrf4 expression becomes undetectable (Grifone et al., 2005). Myf5 expression is correctly initiated but becomes restricted to the caudal region of each somite and is severely impaired in the limb buds of *Six1-/-* and *Six1-/-Six4-/+* mouse mutants (Giordani et al., 2007; Grifone et al., 2007; Grifone et al., 2005). Six1/4 therefore regulate *Myf5* transcription (Giordani et al., 2007; Grifone et al., 2005). There appears to be a collaboration in function between Six1 and Six4 in mouse that is not conserved in *Drosophila* nor is the role of Six1 in myogenesis.

1.5.4 How do Six4 and Six5 function compare with Six1?

I have briefly mentioned how Six4 is involved together with Six1 in the development of the thymus, the kidney and some aspects of neurogenesis and I have explained above the role that Six4 and Six1 perform in the developing somites. The expression pattern of Six4 is similar to that of Six1 except at the early stage of embryonic day 8.5. Six4-deficient mice are born with normal gross appearance and are fertile (Ozaki et al., 2001). *Six4*-deficient embryos show no morphological abnormalities, and the expression patterns of several molecular markers, e.g., myogenin and NeuroD3 (neurogenin1), are normal (Ozaki et al., 2001). *Six4* is thus not essential for mouse embryogenesis and suggests that other members of the Six family seem to compensate for the loss of *Six4* (Ozaki et al., 2001).

Six5 is involved in the development and function of mesodermal tissues and brain (Sato et al., 2002).Heterozygous loss of *Six5* in mice causes cataracts and cardiac conduction disease, and homozygous loss also leads to sterility and decreased testicular mass, reminiscent of DM1 in humans (Sato et al., 2002). There is indeed a strict requirement of *Six5* for both spermatogenic cell survival and spermiogenesis (Sato et al., 2002). Muscle contractile properties, electromyographic insertional activity, and muscle histology are normal in SIX5 deficient mice but leydig cell hyperproliferation, increased intra-testicular testosterone levels and increased FSH levels are observed (Sato et al., 2002). These results support the hypothesis that the reduced SIX5 levels contribute to the male reproductive defects in DM1 (Sato et al., 2002). Studies have demonstrated that a dose-dependent loss of Dm15 (the mouse DMPK homologue) in mice produces a partial DM1 phenotype characterized by decreased development of skeletal muscle force and cardiac conduction disorders

(Sato et al., 2002). The rate and severity of cataract formation, thought to be due to altered ion homeostasis within the lens, is inversely related to Six5 dosage and is temporally progressive (Sato et al., 2002). As ocular cataracts are a characteristic feature of DM1, these results demonstrate that decreased SIX5 transcription is important in the aetiology of DM1 (Sato et al., 2002). This supports the hypothesis that DM1 is a contiguous gene syndrome associated with the partial loss of both DMPK and SIX5 (Sato et al., 2002).

The involvement of *Six5* in muscle development in vertebrates is comparable to that of D-Six4 in *Drosophila.*

1.5.5 How do these expression patterns and functions compare with those of the Six genes in D. melanogaster?

Both in humans and in mouse Six1 and Six4 are coexpressed. Together they are part of the cause for BOR in humans and in mouse they drive the development of somites, the thymus, the kidney and some aspects of neurogenesis. Six5 is also involved in some of these processes. Both in humans and mouse, the Six3/Six6 proteins are expressed primarily in the brain and the visual system and drive the development of these processes. The So/Optix segregation from D-Six4 thus does not appear to have been conserved in mammals. While So and Optix are primarily involved in the development of the fly visual system and D-Six4 in the development of the muscle and gonads, it would seem that there is segregation between the Six1/Six2 and Six4/Six5 proteins from the Six3/Six6 proteins. Additionally, as discussed while Six1/Six2 and Six4/Six5 act as transcriptional activators, Six3/Six6 seem to function as transcriptional repressors. These findings are consistent with the

phylogenetic analysis indicating that Six3/Six6 are more distantly related to the other mammal Six genes.

These differences raise questions about the evolution by gene duplication and divergence of the Six genes. What were the functions of the archetypal *Six* gene? When, how and why did expansion of the gene family occur? How did this correlate with the acquisition of new functions? How have molecular mechanisms diverged? Also, what are the similarities in functional capabilities between the different orthologues? So is almost solely involved in eye development in *Drosophila* and plays no known part in myogenesis. Six1 in mammals on the contrary is very involved in myogenesis and seems to have a much broader role in development of many tissues. Similarly, whereas D-Six4 is principally involved in the patterning of the mesoderm, in mammals, its orthologue Six4, together with being involved in myogenesis, is also involved other processes such as in the development of the trigeminal ganglion. I explained how Six5 in mouse, when knocked out resembled the DM1 phenotype in humans. Similarly in flies, the *D-Six4*^{289} homozygous mutant embryos phenotype have been suggested to echo the DM1 phenotype. How similar are the functional capabilities of these proteins? Optix is sufficient to ectopically drive the expression of an eye and similarly Six3 promotes the formation of ectopic optic vesicle-like structures.

Both in flies and mammals, there are Six/Eya interactions. What are the nature of these interactions, are they necessary, are they sufficient? Are they comparable? Do they involve the same binding sites? Neither Optix nor Six3 interact with Eya. These questions lead to the main title of this PhD: Are the functions of the Six family of transcription factors conserved or have they diverged?

1.6 What was the aim of this PhD?

The aim of this PhD was to assess whether the Six family of transcriptional factors had conserved or diverged functions.

I hypothesised that the Six proteins are functionally distinct but that at least some of these functional differences are conserved between species.

Firstly, I carried out a thorough analysis of the *Drosophila* Six protein sequences in order to complement and refine the criteria defining and classifying Six proteins. These criteria were used as a means to search for Six protein sequences in other species and a phylogeny of the Six family in basal metazoan was constructed. These protein sequence comparisons were then used to identify differences between the three subfamilies and highlight features of putative functional importance and relate back to the question of archetypal and acquired functions correlated with Six family expansion and divergence.

Functional conservation was then tested between the *Drosophila Six* genes. Finally experiments towards the analysis of mammalian *Six* genes in *Drosophila* were also carried out. Below is a summary of each of the chapters in this thesis.

1.7 What will each chapter be about?

Chapter 2 focuses on the characterisation of the Six family. I analysed protein sequence comparisons in order to complement the previous criteria defining and enabling the classification of the Six family.

Chapter 3 focuses on the Six family phylogeny. I searched the databases available for Six proteins in basal metazoan and using the criteria I had established beforehand, I annotated and classified these sequences.

Chapter 4 focuses on features of putative functional conservation. Using the sequence comparisons that I had carried out both in chapter 2 and chapter 3, I identified features of putative functional conservation in the *Drosophila* Six proteins.

Chapter 5 focuses on the conservation and divergence of function of the Six family of transcription factors in *D. melanogaster*. I carried out rescue experiments of aspects of the *D-Six4*^{289} mutant phenotype using the three *D. melanogaster Six* genes as well as *eya.* I showed that while e*ya* is not able to rescue aspects of the *D-Six4*^{289} mutant phenotype, *D-Six4*, *so* and *Optix* show varying capabilities to do so indicating varying levels of functional conservation.

Chapter 6 focuses on the conservation and divergence of function of the Six family of transcription factors in mammals. I attempted to generate molecular constructs containing mammal *Six* genes in order to then inject them into *Drosophila* embryos and assess wether they are able to rescue aspects of the *D-Six4*^{289} mutant phenotype. I report on the various experimental difficulties encountered and on the learning outcomes developed as part of my training as a scientist.

In chapter 7 I discuss the results obtained in each of the results chapter in the light of the current literature available.

Chapter 8 is the materials and methods I used.

Chapter 2

Characterisation of the Six protein family

Characterisation of the Six protein family

2.1 Introduction

Protein sequence comparisons are a means of characterising proteins. Relationships between different species' proteins can be highlighted and conserved features identified. During my PhD, I investigated the functional relationship between Six proteins in *D. melanogaster* and *Mus musculus (M. musculus).* Consequently, the relationship between proteins from the *Drosophila* genus, the insect and mammalian orders are relevant. I compared twenty-four Six protein sequences in eight *Drosophila* species to twelve Six protein sequences in *Anopheles gambiae*, *Apis mellifera*, *M. musculus* and *Homo sapiens*. My aims were to identify criteria defining and classifying the Six protein sequences, thereby contributing to the characterisation of the Six protein family, and understand the relationship between these proteins in these species. A further aim was to identify features of putative functional conservation.

2.1.1 Why characterise the Six protein family?

In 1999, a comprehensive study of the *Six* class homeobox genes was carried out reporting three *Six* genes in *D. melanogaster* and six in mammals (Seo et al., 1999) (Chapter 1). Comparisons of the corresponding protein sequences resulted in a preliminary characterisation of the Six proteins and in their classification into three subfamilies (Seo et al., 1999). Each subfamily, Six1/Six2, Six3/Six6 and Six4/Six5 respectively includes Six proteins orthologous to So, Optix and D-Six4 respectively. According to this classification, proteins within a subgroup are more closely related to one another than they are to the other proteins within the same species (Seo et al.,

1999). No recent comparison of Six protein sequences in insects and mammals has been carried out.

In the eight years since this study a number of genomes have been sequenced including *D. melanogaster* (Adams et al., 2000), *Homo sapiens* (Venter et al., 2001), *M. musculus* (Waterston et al., 2002), *Anopheles gambiae* (Holt et al., 2002), *Drosophila pseudoobscura* (Richards et al., 2005) and *Apis mellifera* (2006). In addition, partial genome sequences of other species are available and Six protein sequences may be obtained from them.

In the light of these available sequences, I questioned whether the original characterisation and classification of the Six protein family (Seo et al., 1999) remain valid. I collected Six protein sequences from several *Drosophila* species, *Apis mellifera* and *Anopheles gambiae* and compared them with the mammalian sequences formerly described by Seo *et al* (Seo et al., 1999). I thereby further characterised the Six proteins by establishing additional criteria to define them and support the previous classification. An understanding of the relationship between these proteins in the different species compared was also attained.

2.1.2 Are there features of putative functional conservation?

In this chapter, I identified conserved features among different protein sequences and suggested that they are of functional importance. Although I was unable to test this functional importance experimentally, I report on the results of two independent studies which in part support experimentally the importance of these features. Furthermore, in following chapters, I will discuss the conservation and divergence of function of the Six family homeodomain transcription factors. In this chapter, I

suggest that their different functional abilities may be attributed to the differences and similarities in the sequence of these conserved features.

2.2. Results

In this chapter I report the collection, alignment and comparisons between species of the Six domain and homeodomain of Six protein sequences.

2.2.1 How did I collect the sequences?

I obtained the *D. melanogaster* sequences of D-Six4, So and Optix and the mammalian sequences of SIX1, SIX2, SIX3, SIX6, SIX4 and SIX5 from the Ensembl Genome Browser (tables 2.2.1.1, 2.2.1.2 and 2.2.1.3).

I used the *D. melanogaster* sequences in a blast search using the BLAST service on Flybase in order to collect the other *Drosophila* sequences as well as the sequences for *Anopheles gambiae* and *Apis mellifera* (Crosby et al., 2007). I found a number of hits for each *Drosophila* species as well as for *Anopheles gambiae and Apis mellifera.* I selected the best match for further comparisons (tables 2.2.1.1, 2.2.1.2 and 2.2.1.3). I initially collected these protein sequences early on during my PhD however, new genomic sequence data subsequently became available and I have included it alongside the previous data. The added tables and figures are numbered similarly to the old data but with and added R for reviewed.

Species	Accession number
D. melanogaster	CG11121
Drosophila simulans	>gnl|Dsim-washu-w501-asm|Contig1.1019
Drosophila yakuba	>gnl|Dyak-washu-assembly_040407|Contig2.89
Drosophila erecta	>gnl|Dere-agencourt-run1028-asm|Contig_73
Drosophila ananassae	>gnl|Dana-agencourt-040714-asm|Contig_796
Drosophila pseudoobscura	XR_groupe8:1..9190824
Drosophila mojavensis	>gnl|Dmov-agencourt-run0811-asm|Contig_1590
Drosophila virilis	> gnl|Dvir-agencourt-run1029-asm|Contig_476
Anopheles gambiae	>gi|19612305|gb|AAAB01008986|AAAB01008986
Apis mellifera	>gi|63051463|gb|AADG05001993|AADG05001993
M. musculus	ENSMUSG00000051367 (six1) ENSMUSG00000024134 (six2)
Homo sapiens	ENSG00000126778 (six1) ENSG00000170577 (six2)

Table 2.2.1.1: Accession numbers of the *D. melanogaster* So protein sequence and of the orthologous protein sequences compared.

Species	Accession number
Drosophila yakuba	>gnl|dyak|FBpp0264134
Drosophila erecta	>gnl|dere|FBpp0141823
Drosophila ananassae	>gnl|dana|FBpp0116147
Drosophila mojavensis	>gnl|dmoj|FBpp0167976
Drosophila virilis	>gnl|dvir|FBpp0236200
Anopheles gambiae	>gi|157013450|gb|EAA00668.4| AGAP011695-PA
Apis mellifera	>gi|110756670|ref|XP_396811.3|

Table 2.2.1.1R: New accession numbers of the orthologous protein sequences compared with *D. melanogaster* So.

Species	Accession number
D. melanogaster	CG18455
Drosophila simulans	> gi|62998562|gb|CM000165|CM000165
Drosophila yakuba	> gi|49845347|gb|AAEU01000660|AAEU01000660
Drosophila erecta	>gnl|Dere|Contig_194
Drosophila ananassae	>gnl|Dana|Contig_683
Drosophila pseudoobscura	XR_groupe8:1..9190824
Drosophila mojavensis	>gnl|DmoJ|Contig_1323
Drosophila virilis	> gnl|Dvir|Contig_536
Anopheles gambiae	>gi|19611829|gb|AAAB01008844|AAAB01008844
Apis mellifera	>gi|63052512|gb|AADG05000944|AADG05000944
M. musculus	ENSMUSG00000038805 (six3) ENSMUSG00000021099 (six6)
Homo sapiens	ENSG00000138083 (six3) ENSG00000184302 (six6)

Table 2.2.1.2: Accession numbers of the *D. melanogaster* Optix protein sequence and of the orthologous proteins sequences compared.

Species	Accession number
Drosophila mojavensis	>gnl|dmoj|FBpp0168564

Table 2.2.1.2R: New accession number of the orthologous protein sequence compared with *D. melanogaster* Optix.

Species	Accession number
D. melanogaster	CG3871
Drosophila simulans	>gnl|Dsim-washu-w501-asm|Contig1.1019
Drosophila yakuba	>gnl|Dyak-washu-assembly_040407|Contig135.3
Drosophila erecta	>gnl|Dere-agencourt-run1028-asm|Contig_513
Drosophila ananassae	>gnl|Dana-agencourt-040714-asm|Contig_204
Drosophila pseudoobscura	XR_groupe8:1..9190824
Drosophila mojavensis	>gnl|Dmov-agencourt-run0811-asm|Contig_2976
Drosophila virilis	> gnl|Dvir-agencourt-run1029-asm|Contig_2320
Anopheles gambiae	>gi|19611762|gb|AAAB01008816|AAAB01008816
Apis mellifera	>gi|63053020|gb|AADG05000436|AADG05000436
M. musculus	ENSMUSG00000034460 (six4) ENSMUSG00000040841 (six5)
Homo sapiens	ENSG00000100625 (six4) ENSG00000177045 (six5)

Table 2.2.1.3: Accession numbers of the *D. melanogaster* D-Six4 protein sequence and of the orthologous protein sequences compared.

Species	Accession number
Drosophila simulans	>gnl|dsim|FBpp0210563
Drosophila yakuba	>gnl|dyak|FBpp0267742
Drosophila erecta	>gnl|dere|FBpp0131829
Drosophila ananassae	>gnl|dana|FBpp0127303
Drosophila pseudoobscura	>gi|54641765|gb|EAL30515.1| GA17741-PA
Drosophila mojavensis	>gnl|dmoj|FBpp0161104
Drosophila virilis	>gnl|dvir|FBpp0228183
Apis mellifera	>gi|110749482|ref|XP_001120698.1|

Table 2.2.1.3R: New accession numbers of the orthologous protein sequences compared with *D. melanogaster* D-Six4.

2.2.2 How did I align the sequences?

I aligned the sequences using the European Bioinformatics Institute sequence alignment programme, ClustalW (Thompson et al., 1994). ClustalW is a general purpose multiple sequence alignment program that can be used to compare protein sequences. It produces multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences, and lines them up thus highlighting identities and differences. In figures and tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3, I abbreviated the insect species' names to three letter codes (table 2.2.2.1). In figures 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3, mouse Six proteins were preceded by the letter 'm' (msix1, msix2 etc.) and human Six proteins were preceded by the letter 'h' (hsix1, hsix2 etc) and in tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3, I further abbreviated them to M1, M2 and H1, H2 etc.

Three letter code	Species
SIM	*Drosophila simulans*
YAK	*Drosophila yakuba*
ERE	*Drosophila erecta*
ANA	*Drosophila ananassae*
PSE	*Drosophila pseudoobscura*
MOJ	*Drosophila mojavensis*
VIR	*Drosophila virilis*
ANO	*Anopheles gambiae*
API	*Apis mellifera*

Table 2.2.2.1: This table shows the three letter codes used in the sequence comparisons.

2.2.3 What did the comparisons of insect and mammalian Six domains show?

I initially carried out these comparisons early on during my PhD and the analysis of them is described below. However, I subsequently carried out a comparison of the new genomic sequence data available and I have included it alongside the previous data. The comparisons I carried out of the Six domains of the Six protein sequences of *D. melanogaster* and their orthologues in other *Drosophila* species (figures 2.2.3.1, 2.2.3.2 and 2.2.3.3) show 100% identity between Optix and its orthologues and between most orthologues of D-Six4 and So respectively (tables 2.2.3.1, 2.2.3.2 and 2.2.3.3). The D-Six4 orthologous protein sequences from *Drosophila yakuba, Drosophila mojavensis* and *Drosophila virilis* show a lower percentage of identity with *D. melanogaster* D-Six4 protein (97%, 98% and 97% respectively) and the orthologous So protein sequence from *Drosophila yakuba* also shows a lower percentage of identity with *D. melanogaster* So protein (91%) (tables 2.2.3.1 and 2.2.3.2). These lower percentages were observed initially but in the light of the new data available, it appears that the Six domains of the three Six protein sequences (D-Six4, So and Optix) of *D. melanogaster* show 100% identity with their respective orthologues in other *Drosophila* species (tables 2.2.3.1R-2.2.3.3R and figures 2.2.3.1R-2.2.3.3R). In the original comparison I carried out, The *Drosophila mojavensis* orthologue of Optix showed an insertion of twenty four amino acids (figure 2.2.3.3). Insertions were also noticeable in *Drosophila erecta, yakuba and ananasse* orthologues of So (figure 2.2.3.2). These insertions were absent in subsequent sequencing experiments (figures 2.2.3.2R and 2.2.3.3R) and were consequently an artefact, symptomatic of the fact that I had obtained these sequences

from preliminary genome sequencing projects. I obtained the new genomic sequence information subsequently.

The comparisons I carried out of the Six domains of the Six protein sequences of *D. melanogaster* and their othologues in *Anopheles gambiae* and *Apis mellifera* (figures 2.2.3.1, 2.2.3.2 and 2.2.3.3) show that in these latter two species the highest percentage of identity is between So and its orthologues, (91% and 81% respectively for *Anopheles gambiae* and *Apis mellifera*) (tables 2.2.3.1, 2.2.3.2 and 2.2.3.3). It decreases considerably between both Optix and D-Six4 and their orthologues (80% and 77% in *Anopheles gambiae* respectively and 78% and 68% in *Apis mellifera*) (tables 2.2.3.1, 2.2.3.2 and 2.2.3.3).

I verified the previous comparisons between *D. melanogaster* and mammalian Six protein sequences (Seo et al., 1999). Consistent with previous results, the comparisons I carried out show 84% identity between So and mammal Six2 (Seo et al., 1999) (table 2.2.3.2). I found a one percentage difference for the percentages of identity between D-Six4 and mammal Six4 (56% compared to 57% (Seo et al., 1999)) and Optix and mammal Six3 (76% compared to 77% (Seo et al., 1999)) (table 2.2.3.1 and 2.2.3.3). I considered these differences insignificant. Additionally, I completed previous comparisons and I showed that So, Optix and D-Six4 display 85 %, 73% and 59% identity with mammal Six1, Six6 and Six5 respectively (tables 2.2.3.1, 2.2.3.2 and 2.2.3.3).

From the comparisons that I carried out, I report that the Six domains of the two mammalian paralogues in each subclass show varying percentages of identity (88%, 65% and 95% between Six3/Six6, Six4/Six5 and Six1/Six2 respectively) (tables 2.2.3.1, 2.2.3.2 and 2.2.3.3). I also report that the Six domains of the two mammalian

orthologues in each subclass are identical except for mouse Six5 and human Six5 (99%) (table 2.2.3.1).

I explained in chapter 1 that Six4-deficient mice are born with normal gross appearance and Six4-deficient embryos show no morphological abnormalities. It has been suggested that other members of the Six family compensate for the loss of Six4. Six1 and Six4 have similar expression patterns. It has been proposed that Six1 functions instead of Six4 in Six4 deficient mice (Giordani et al., 2007). In light of this, I added mouse Six1 sequences to the Six4/Six5 comparisons in order to see if sequence similarities would support my hypothesis. Interestingly, the protein sequence of the Six domain of mouse Six4 shares a higher percentage of identity with the protein sequence of the Six domain of mouse Six1 (72%) than it does with the protein sequence of the Six domain of mouse Six5 (65%) (table 2.2.3.1).

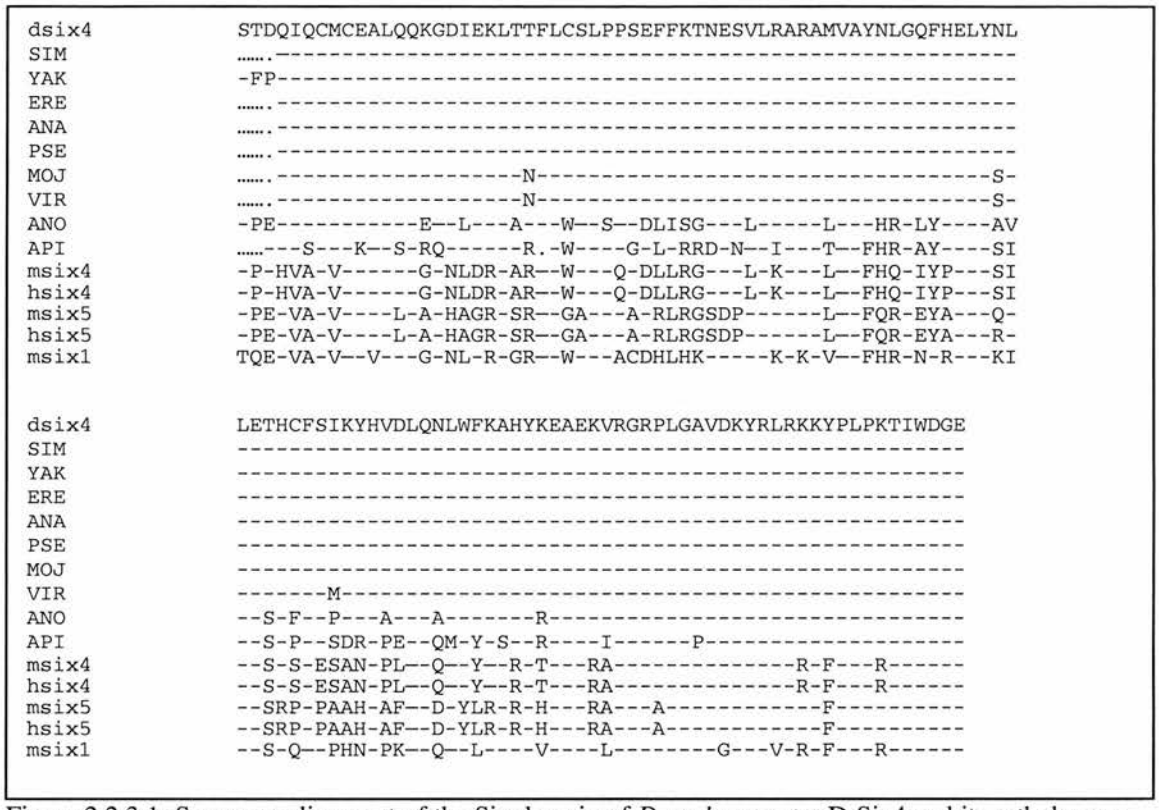


Figure 2.2.3.1: Sequence alignment of the Six domain of *D. melanogaster* D-Six4 and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

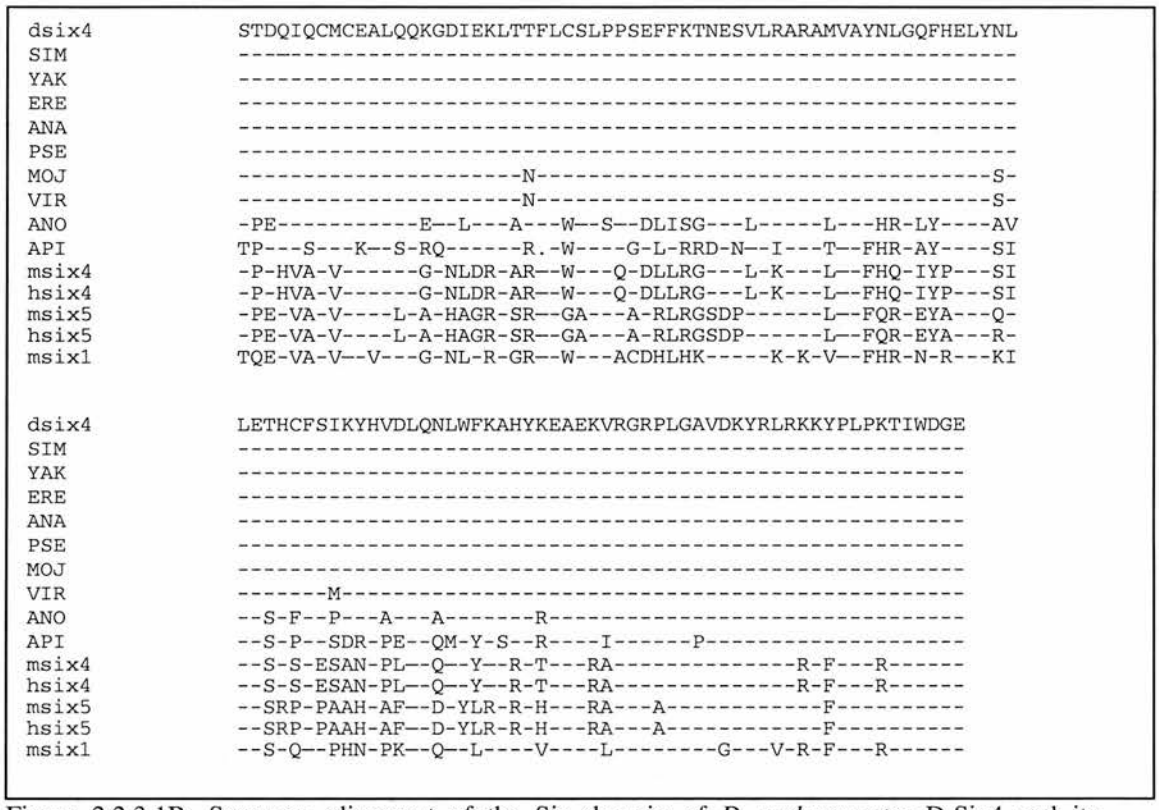


Figure 2.2.3.1R: Sequence alignment of the Six domain of *D. melanogaster* D-Six4 and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown.

SIX	**Dsix4**	Sim	Yak	Ere	Ana	Pse	Moj	Vir	Ano	Api	M4	H4	M5	H5	M1
Dsix4	**100**														
Sim	**100**	**100**													
Yak	**97**	100	**100**												
Ere	**100**	100	100	**100**											
Ana	**100**	100	100	100	**100**										
Pse	**100**	100	100	100	100	**100**									
Moj	**98**	98	98	98	98	98	**100**								
Vir	**97**	97	97	97	97	97	99	**100**							
Ano	**77**	78	76	78	78	78	77	77	**100**						
Api	**68**	67	67	67	67	67	68	68	68	**100**					
M4	**59**	59	57	59	59	59	60	60	68	63	**100**				
H4	**59**	59	57	59	59	59	60	60	68	63	100	**100**			
M5	**56**	57	56	57	57	57	57	57	60	55	65	65	**100**		
H5	**56**	57	56	57	57	57	57	57	60	55	65	65	99	**100**	
M1	**59**	61	59	61	61	61	61	61	63	60	72	72	57	57	**100**

Table 2.2.3.1: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.1. The cells colored in blue relate to the *Drosophil*a genus. The cells colored in green relate to the other insects. The cells colored in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

SIX	**Dsix4**	Sim	Yak	Ere	Ana	Pse	Moj	Vir	Ano	Api	M4	H4	M5	H5	M1
Dsix4	**100**														
Sim	**100**	**100**													
Yak	**100**	100	**100**												
Ere	**100**	100	100	**100**											
Ana	**100**	100	100	100	**100**										
Pse	**100**	100	100	100	100	**100**									
Moj	**98**	98	98	98	98	98	**100**								
Vir	**97**	97	97	97	97	97	99	**100**							
Ano	**77**	77	77	77	77	77	77	77	**100**						
Api	**66**	66	66	66	66	66	65	65	67	**100**					
M4	**59**	59	59	59	59	59	60	60	68	63	**100**				
H4	**59**	59	59	59	59	59	60	60	68	63	100	**100**			
M5	**56**	57	57	57	57	57	57	57	60	55	65	65	**100**		
H5	**56**	57	57	57	57	57	57	57	60	55	65	65	99	**100**	
M1	**59**	61	61	61	61	61	61	61	63	60	72	72	57	57	**100**

Table 2.2.3.1R: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.1R. The cells colored in blue relate to the *Drosophil*a genus. The cells colored in green relate to the other insects. The cells colored in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

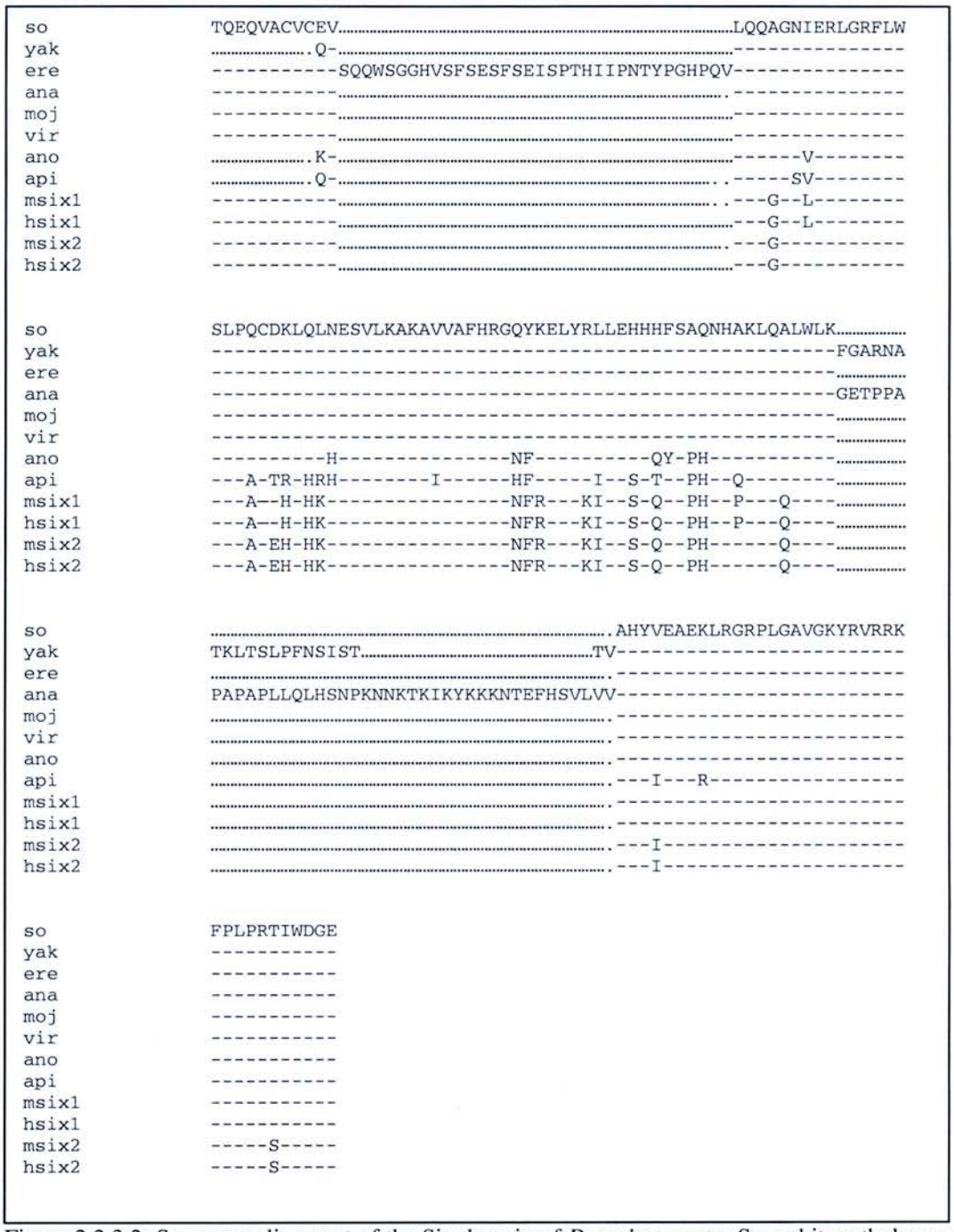


Figure 2.2.3.2: Sequence alignment of the Six domain of *D. melanogaster* So and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

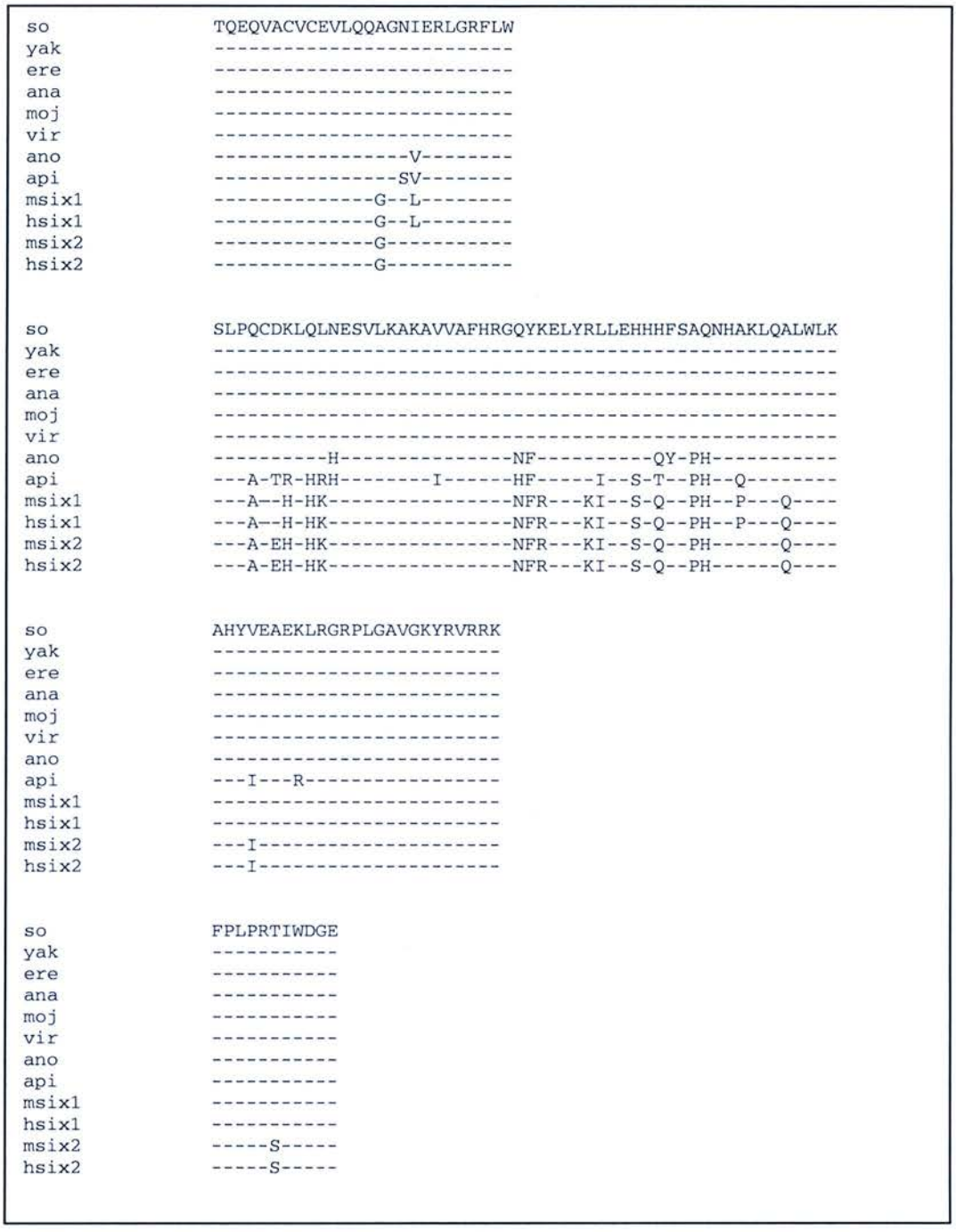


Figure 2.2.3.2R: Sequence alignment of the Six domain of *D. melanogaster* So and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown.

Six	So	Yak	Ere	Ana	Moj	Vir	Ano	Api	M1	H1	M2	H2
So	100											
Yak	91	100										
Ere	100	83	100									
Ana	100	85	77	100								
Moj	100	91	100	100	100							
Vir	100	91	100	100	100	100						
Ano	91	91	91	91	91	91	100					
Api	81	82	82	81	81	81	85	100				
M1	85	76	85	85	85	85	85	84	100			
H1	85	76	85	85	85	85	85	84	100	100		
M2	84	75	84	84	84	84	84	84	95	95	100	
H2	84	75	84	84	84	84	84	84	95	95	100	100

Table 2.2.3.2: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.2. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

Six	So	Yak	Ere	Ana	Moj	Vir	Ano	Api	M1	H1	M2	H2
So	100											
Yak	100	100										
Ere	100	100	100									
Ana	100	100	100	100								
Moj	100	100	100	100	100							
Vir	100	100	100	100	100	100						
Ano	93	93	93	93	93	93	100					
Api	83	83	83	83	83	83	87	100				
M1	85	85	85	85	85	85	85	84	100			
H1	85	85	85	85	85	85	85	84	100	100		
M2	84	84	84	84	84	84	84	84	95	95	100	
H2	84	84	84	84	84	84	84	84	95	95	100	100

Table 2.2.3.2R: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.2R. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

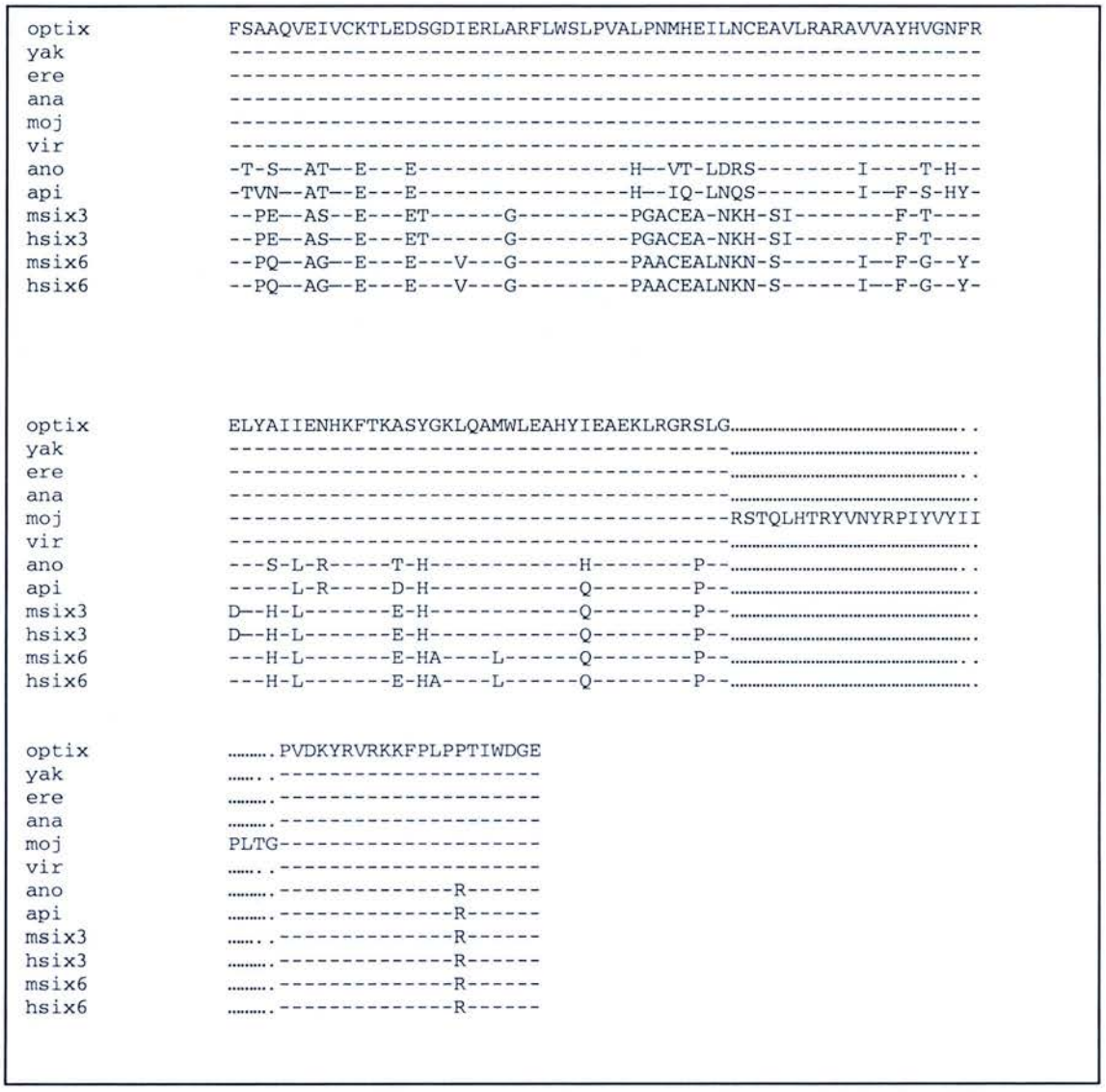


Figure 2.2.3.3: Sequence alignment of the Six domain of *D. melanogaster* Optix and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to Optix are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

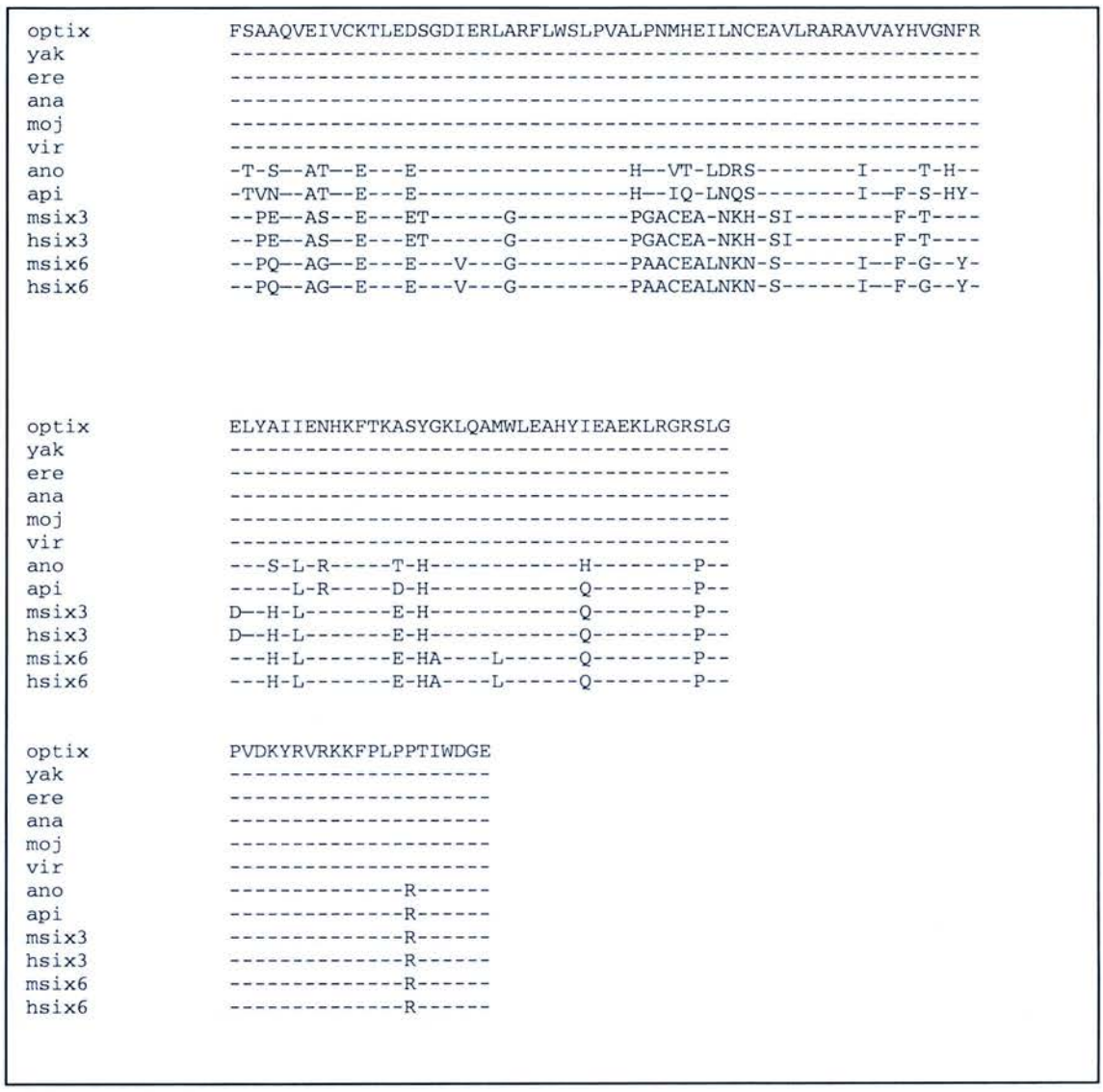


Figure 2.2.3.3R: Sequence alignment of the Six domain of *D. melanogaster* Optix and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to Optix are indicated by hyphens and the differing amino acids are shown.

Six	Optix	Yak	Ere	Ana	Moj	Vir	Ano	Api	M3	H3	M6	H6
Optix	100											
Yak	100	100										
Ere	100	100	100									
Ana	100	100	100	100								
Moj	100	100	100	100	100							
Vir	100	100	100	100	100	100						
Ano	80	80	80	80	80	80	100					
Api	78	78	78	78	78	78	90	100				
M3	76	76	76	76	76	76	78	79	100			
H3	76	76	76	76	76	76	78	79	100	100		
M6	73	73	73	73	73	73	78	81	88	88	100	
H6	73	73	73	73	73	73	78	81	88	88	100	100

Table 2.2.3.3: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.3. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

Six	Optix	Yak	Ere	Ana	Moj	Vir	Ano	Api	M3	H3	M6	H6
Optix	100											
Yak	100	100										
Ere	100	100	100									
Ana	100	100	100	100								
Moj	100	100	100	100	100							
Vir	100	100	100	100	100	100						
Ano	80	80	80	80	80	80	100					
Api	78	78	78	78	78	78	90	100				
M3	76	76	76	76	76	76	78	79	100			
H3	76	76	76	76	76	76	78	79	100	100		
M6	73	73	73	73	73	73	78	81	88	88	100	
H6	73	73	73	73	73	73	78	81	88	88	100	100

Table 2.2.3.3R: Percentages of identity between the Six domains of the species aligned in figure 2.2.3.3R. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

2.2.4 What did the comparisons of insect and mammalian homeodomains show?

Similarly to the analysis carried out for the Six domain, I initially carried out these comparisons early on during my PhD and their analysis is described below. The subsequent new data (figure 2.2.4R) is included alongside the previous data. The added figures are numbered similarly to the old data but with an added R for reviewed. The comparisons I carried out of the homeodomains of the Six protein sequences of *D. melanogaster* and their orthologues in other *Drosophila* species (figures 2.2.4.1, 2.2.4.2 and 2.2.4.3) show 100% identity between Optix and D-Six4 and their respective orthologues and between So and most of its orthologues (original data) (tables 2.2.4.1, 2.2.4.2 and 2.2.4.3). Initially, the So orthologous protein sequences from *Drosophila erecta, mojavensis and virilis* showed a lower percentage of identity to the *D. melanogaster* So protein (94%, 98% and 80% respectively) (table 2.2.4.2). These lower percentages were observed initially but in the light of the new data available it appears that similarly to the homeodomains of Optix and D-Six4 and their respective orthologues in other *Drosophila* species, the sequence of the homeodmain of So is 100% identical to that of the sequences of the homeodomain of the other *Drosophila* species compared. Initially, the So orthologous protein sequence from *Drosophila virilis* also exhibited an insertion of twenty two amino acids (figure 2.2.4.2). This insertion was absent from subsequent sequencing experiments indicating that it was an artefact of the original sequencing.

The comparisons I carried out of the homeodomains of the Six protein sequences of *D. melanogaster* and their orthologues in *Anopheles gambiae* (figures 2.2.4.1, 2.2.4.2 and 2.2.4.3) show 96 % identity for each Six protein. The *Apis*

mellifera orthologues show 95%, 93% and 78% identity with Optix, So and D-Six4 respectively (tables 2.2.4.1, 2.2.4.2 and 2.2.4.3).

Consistent with previous results, the comparisons I carried out show 95% identity between So and mammal Six2 (Seo et al., 1999) (table 2.2.4.2). I again found a one or two percentage difference for the percentages of identity between D-Six4 and mammal Six4 (81% compared to 82% (Seo et al., 1999) respectively) and Optix and mammal Six3 (95% compared 97% (Seo et al., 1999) respectively) (table 2.2.4.1 and 2.2.4.3). Again, I considered these differences insignificant. Additionally, I showed that So, Optix and D-Six4 display 93 %, 93% and 83% identity with mammal Six1, Six6 and Six5 respectively (tables 2.2.4.1, 2.2.4.2 and 2.2.4.3).

From the comparisons I carried out, I report that the protein sequence of the homeodomain of mouse Six5 shares a higher percentage of identity with the protein sequence of the homeodomain of *D. melanogaster* D-Six4 (83%) than it does with the protein sequence of the homeodomain of mouse Six4 (80%) (table 2.2.4.1). Whether this has any significance remains to be determined.

I also report that the protein sequence of the homeodomains of the mammalian orthologues for each Six protein is identical (tables 2.2.4.1, 2.2.4.2 and 2.2.4.3) and that the homeodomains of the two paralogues in each subclass show varying percentages of identity (98%, 80% and 98% between Six3/Six6, Six4/Six5 and Six1/Six2 respectively) (tables 2.2.4.1, 2.2.4.2 and 2.2.4.3).

```
dsix4           WDGEETVYCFKEKSRNALKDCYLTNRYPTPDEKKTLAKKTGLTLTQVSNWFKNRRQRDRTP
sim             ------------------------------------------------------------
yak             ------------------------------------------------------------
ere             ------------------------------------------------------------
ana             ------------------------------------------------------------
pse             ------------------------------------------------------------
moj             ------------------------------------------------------------
vir             ------------------------------------------------------------
ano             ----------------------TR--------------------------------------
api             -----V------R------Q--MRTK--LSE---N---E--------A-------------
msix4           ------------------EL-KQ----S-A--RH---I---S----------------N-
hsix4           ------------------EL-KQ----S-A--RH---I---S----------------N-
msix5           ------------R--A---A--RG---------RR--TL---S-----------------.
hsix5           ------------R--A---A--RG---------RR--TL---S-----------------.
msix1           ------S--------GV-REW-AH-P--S-R--RE--EA----T---------------AA
```

Figure 2.2.4.1: Sequence alignment of the homeodomain of *D. melanogaster* D-Six4 and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

Homeo	**Dsix4**	Sim	Yak	Ere	Ana	Pse	Moj	Vir	Ano	Api	M4	H4	M5	H5	M1
Dsix4	**100**														
Sim	**100**	**100**													
Yak	**100**	100	**100**												
Ere	**100**	100	100	**100**											
Ana	**100**	100	100	100	**100**										
Pse	**100**	100	100	100	100	**100**									
Moj	**100**	100	100	100	100	100	**100**								
Vir	**100**	100	100	100	100	100	100	**100**							
Ano	**96**	96	96	96	96	96	96	96	**100**						
Api	**78**	78	78	78	78	78	78	78	80	**100**					
M4	**81**	81	81	81	81	81	81	81	81	72	**100**				
H4	**81**	81	81	81	81	81	81	81	81	72	100	**100**			
M5	**83**	83	83	83	83	83	83	83	83	73	80	80	**100**		
H5	**83**	83	83	83	83	83	83	83	83	73	80	80	100	**100**	
M1	**70**	70	70	70	70	70	70	70	70	62	73	73	70	70	**100**

Table 2.2.4.1: Percentages of identity between the homeodomains of the species aligned in figure 2.2.4.1. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

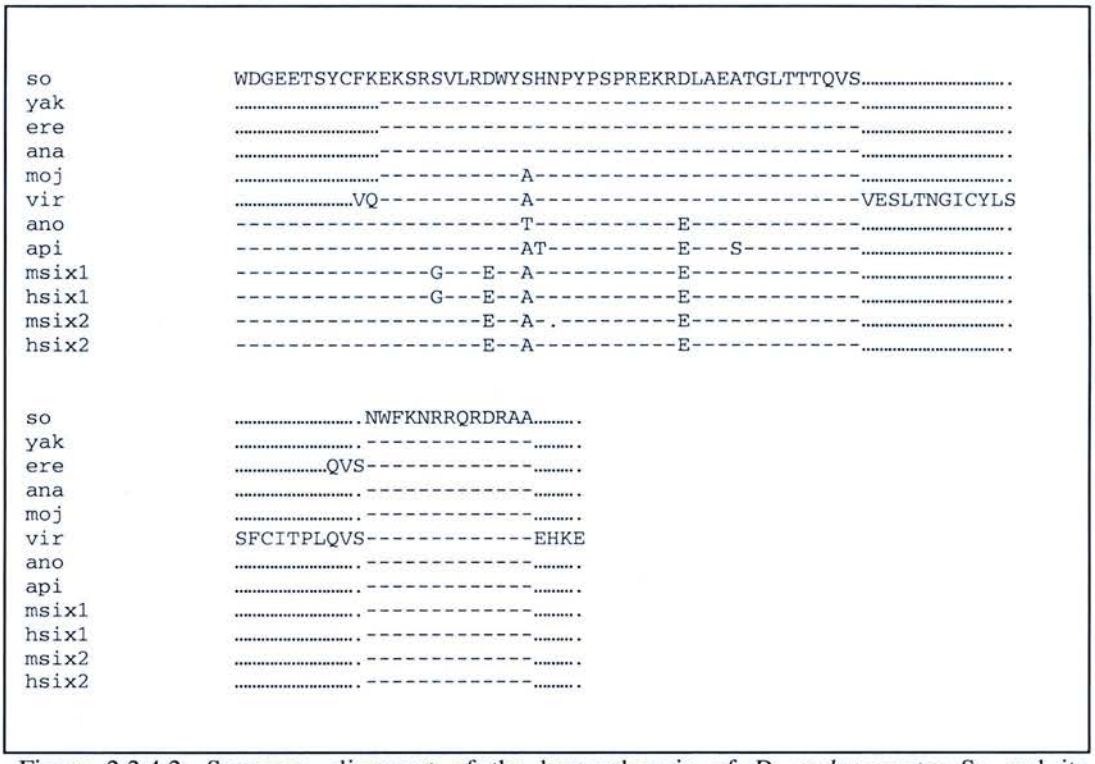


Figure 2.2.4.2: Sequence alignment of the homeodomain of *D. melanogaster* So and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

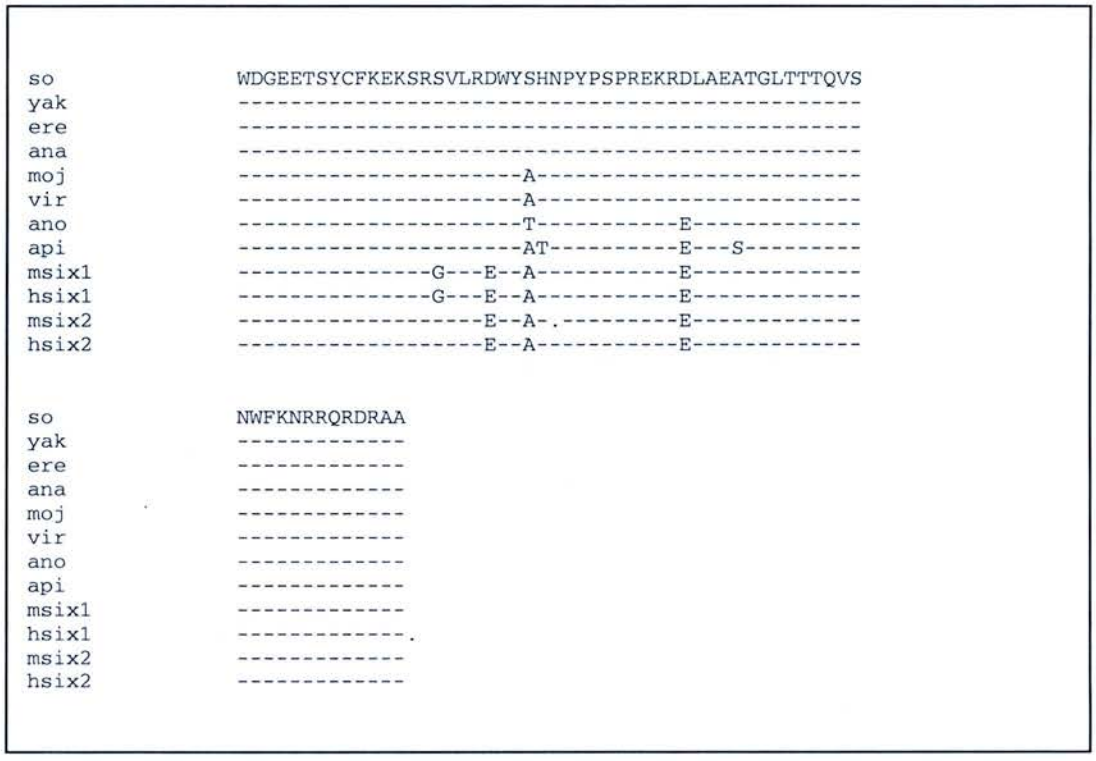


Figure 2.2.4.2R: Sequence alignment of the homeodomain of *D. melanogaster* So and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown.

Homeo	So	Yak	Ere	Ana	Moj	Vir	Ano	Api	M1	H1	M2	H2
So	100											
Yak	100	100										
Ere	94	100	100									
Ana	100	100	100	100								
Moj	98	98	98	98	100							
Vir	80	98	98	98	100	100						
Ano	96	96	90	96	96	78	100					
Api	93	92	86	92	94	77	95	100				
M1	93	92	86	92	94	77	95	93	100			
H1	93	92	86	92	94	77	95	93	100	100		
M2	95	92	86	92	94	78	96	95	98	98	100	
H2	95	94	88	94	96	78	96	95	98	98	100	100

Table 2.2.4.2: Percentages of identity between the homeodomains of the species aligned in figure 2.2.4.2. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

Homeo	So	Yak	Ere	Ana	Moj	Vir	Ano	Api	M1	H1	M2	H2
So	100											
Yak	100	100										
Ere	100	100	100									
Ana	100	100	100	100								
Moj	98	98	98	98	100							
Vir	98	98	98	98	100	100						
Ano	96	96	96	96	96	96	100					
Api	93	93	93	93	95	95	93	100				
M1	93	93	93	93	94	94	95	93	100			
H1	93	93	93	93	94	93	95	93	100	100		
M2	95	95	95	95	94	93	96	95	98	98	100	
H2	95	95	95	95	94	93	96	95	98	98	100	100

Table 2.2.4.2R: Percentages of identity between the homeodomains of the species aligned in figure 2.2.4.2R. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in dark pink show the percentages among the mammal class.

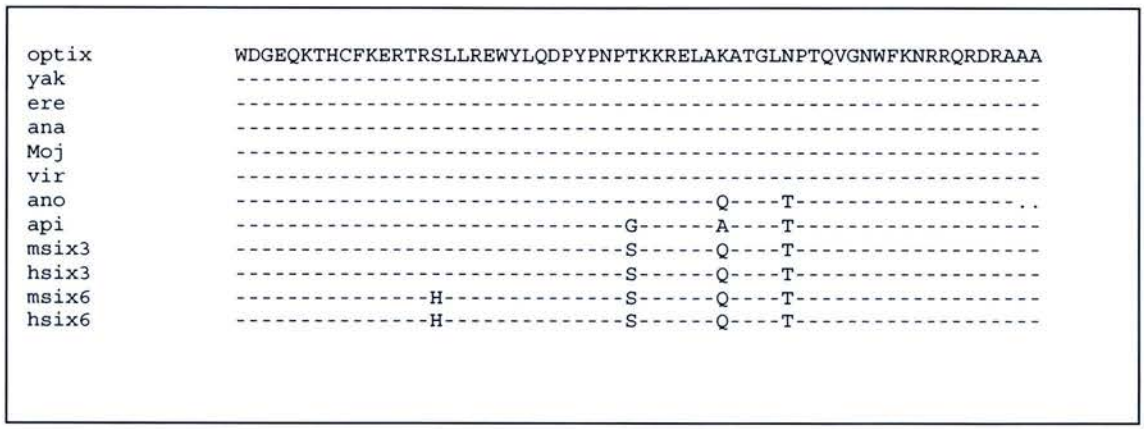


Figure 2.2.4.3: Sequence alignment of the homeodomain of *D. melanogaster* Optix and its orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to Optix are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

Homeo	**Optix**	Yak	Ere	Ana	Moj	Vir	Ano	Api	M3	H3	M6	H6
Optix	**100**											
Yak	**100**	**100**										
Ere	**100**	100	**100**									
Ana	**100**	100	100	**100**								
Moj	**100**	100	100	100	**100**							
Vir	**100**	100	100	100	100	**100**						
Ano	**96**	96	96	96	96	96	**100**					
Api	**95**	95	95	95	95	95	96	**100**				
M3	**95**	95	95	95	95	95	98	96	**100**			
H3	**95**	95	95	95	95	95	98	96	100	**100**		
M6	**93**	93	93	93	93	93	96	95	98	98	**100**	
H6	**93**	93	93	93	93	93	96	95	98	98	100	**100**

Table 2.2.4.3: Percentages of identity between the homeodomains of the species aligned in figure 2.2.4.3. The cells coloured in blue relate to the *Drosophila* genus. The cells coloured in green relate to the other insects. The cells coloured in pink relate to the mammal class and the cells in purple show the percentages among the mammal class.

2.3. Discussion

In this chapter, I collected, aligned and compared the Six domains and homeodomains of Six protein sequences of eight *Drosophila* species, *Anopheles gambiae*, *Apis mellifera*, *M. musculus* and *Homo sapiens*. I thereby characterised the Six protein family by providing additional criteria defining and supporting the 1999 classification. Relationships between these proteins in the insect and mammalian orders have been further understood and I identified features of putative functional conservation. I discuss these in turn below.

2.3.1 How did I characterise the Six protein family?

I showed that the percentages of sequence identities between the Six domains and the homeodomains of the individual *D. melanogaster* Six proteins relative to their respective homologues in the *Drosophila* genus range from 97% to 100% (tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3). I also showed that the percentages of identity range from 56% to 96% in the species compared from the insect and mammal orders. Sequences from the Six4/Six5 subfamily consistently show lower percentages of identity (tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3). I propose that these percentages indicate conservation of these sequences through evolutionary time since the common ancestor of insects and mammals. Consistent with previous results, I report that the homeodomains of the sequences I compared are generally more conserved than the Six domains that show more variability (Seo et al., 1999).

Previous sequence comparisons led to the identification of a tetrapeptide sequence (CFKE) conserved in the Six protein class homeodomains (Seo et al., 1999). This tetrapeptide was defined as a criterion of the Six class homeodomains (Seo et al., 1999). Consistent with this, all the homeodomains of the Six sequences I compared display this same unique peptide sequence (figure 2.3.1.1). Initially, the sequence orthologous to So in *Drosophila virilis* displayed (VQ) instead of (FK) but this was shown to be an artefact of preliminary data. Previously, a consensus sequence for homeodomains was derived from the comparisons of 346 homeodomains of 55 species ranging from *Schizosccaharomyces pombe* to humans. This led to the identification of thirteen amino acid residues very highly conserved (figure 2.3.1.2) (Duboule 1994). Twelve of these were shown to be invariant among the Six class members (Duboule, 1994; Seo et al., 1999). Consistent with this, I report that these twelve amino acids remain invariant in the sequences I compared in this study. Previously, consensus sequences for each of the three subfamilies were derived (Seo et al., 1999). I compared these consensus sequences with the consensus sequence derived from the 346 homeodomains previously compared and I show that twelve amino acids are identical across all four consensus sequences (figure 2.3.1.3) (Duboule, 1994; Seo et al., 1999). These remain identical in all the sequences I compared in this study.

Previous studies have defined the Six domain as a stretch of 116 amino acids upstream of the homeodomain (Cheyette et al., 1994) (chapter 1). From the alignments I carried out in this chapter, I identified a conserved stretch of ten amino acids upstream of the homeodomain with variability according to the family classification ($LG^{P}/_{A}V^{D}/_{G}KYR^{V}/_{L}R$) (figure 2.3.1.4). I propose this decapeptide as a

defining criterion of the Six domain. This decapeptide is conserved in all the species I compared in this study (figure 2.3.1.4). I also verified that it was conserved in the species that were previously compared (Seo et al., 1999). These include chicken, zebrafish, shark, frog, medaka and lamprey (Seo et al., 1999). Their Six domain sequences display this newly identified decapeptide (Seo et al., 1999). This supports my proposal of this decapeptide as a defining feature of the Six class Six domain.

so	WDGEETSYC**FKE**	optix	WDGEQKTHC**FKE**	dsix4	WDGEETVYC**FKE**
yak	----------	yak	----------	sim	----------
ere	----------	ere	----------	yak	----------
ana	----------	ana	----------	ere	----------
moj	----------	Moj	----------	ana	----------
vir	----------	vir	----------	pse	----------
ano	----------	ano	----------	moj	----------
api	----------	api	----------	vir	----------
msix1	----------	msix3	----------	ano	----------
hsix1	----------	hsix3	----------	api	-----V------
msix2	----------	msix6	----------	msix4	----------
hsix2	----------	hsix6	----------	hsix4	----------
				msix5	----------
				hsix5	----------
				msix1	------S-----

Figure 2.3.1.1: First 12 amino acids of the homeodomains of So, Optix and D-Six4 and their orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. The homeodomain characteristic tetrapeptide CFKE is shown in bold. Sequence identities relative to each Six protein are indicated by hyphens and the differing amino acids are shown.

1 RRRK*R*TAYTRYQLLE**L**EKE**F**HFNRY*L*TRRRRIELAHSLN*L*TERQ*V*K*I***WF**Q**N**R**R**MKW***K***KEN 60

Figure 2.3.1.2: Consensus sequence derived from the comparisons of 346 homeodomains. Amino acids in bold and/or in italics are the seven and six extremely and highly conserved amino acids respectively(Duboule, 1994). These remain invariant in all the sequences compared in this study.

Consensus six3	.GEQKTHCFKERTRL**L**REWYLQDPYPNPSKKRELAQATGLTPTQ*V*GN**WF**K**N**R**R**QRDRAAA	59
Consensus six2	.GEETSYCFKEKSRV**L**REWYAHNPYPSPREKRELAEATGLTTTQ*V*SN**WF**K**N**R**R**QRDRAAE	59
Consensus six4	.GEETVYCFKESRNA**L**K.....YNRYP.P.EKR.LAK.TGLSLTQ*V*SN**WF**K**N**R**R**QRDRP..	49
Consensus HD	RRRK*R*TAYTRYQLLE**L**EKE**F**HFNRY*L*TRRRRIELAHSLN*L*TERQ*V*K*I***WF**Q**N**R**R**MKW***K***KEN	60

Figure 2.3.1.3: Alignment of the consensus sequences of the homeodomains of the three families of Six class proteins and the consensus sequence in figure 2.3.1.2. Amino acids in bold and/or in italics are the seven and six extremely and highly conserved amino acids respectively. Amino acids identical in all four consensus sequences are shown in bigger font(Duboule, 1994; Seo et al., 1999).

so	LGAVGKYRVR	optix	LGPVDKYRVR	dsix4	LGAVDKYRLR
yak	----------	yak	----------	SIM	----------
ere	----------	ere	----------	YAK	----------
ana	----------	ana	----------	ERE	----------
moj	----------	moj	----------	ANA	----------
vir	----------	vir	----------	PSE	----------
ano	----------	ano	----------	MOJ	----------
api	----------	api	----------	VIR	----------
msix1	----------	msix3	----------	ANO	----------
hsix1	----------	hsix3	----------	API	--P-------
msix2	----------	msix6	----------	msix4	----------
hsix2	----------	hsix6	----------	hsix4	----------
				msix5	----------
				hsix5	----------

Figure 2.3.1.4: 19 amino acids upstream of the homeodomain, the Six domain characteristic decapepetide of So, Optix and D-Six4 and their orthologues in the *Drosophila* genus, *Anopheles gambiae, Apis mellifera, M. musculus* and *Homo sapiens*. Sequence identities relative to each Six protein are indicated by hyphens and the differing amino acids are shown.

2.3.2 What is the functional relevance of these conserved features?

Previous work established the tetrapeptide (CFKE) as a defining feature of the Six family homeodomain transcription factors on the sole basis of its conservation in protein sequences in the species compared (Seo et al., 1999). In a similar way, I propose in this study to define the Six family Six domain more completely than was previously done by suggesting that the decapepetide ($LG^{P}/_{A}V^{D}/_{G}KYR^{V}/_{L}R$) is a criterion of identification. Previously, the Six domain had simply been defined as the 116 amino acids upstream of the homeodomain (Cheyette et al., 1994). Whilst I was unable during this PhD to corroborate experimentally the functional importance of this decapeptide thereby supporting my proposal, I report here on two independent studies which in part support experimentally the importance of this decapeptide.

2.3.2.1 What available experimental evidence supports the importance of the proposed decapeptide?

In 2005, Kenyon *et al* show that the dominant mutant allele SO^{D} ,which leads to smaller eye discs, is the result of a single amino acid change within the Six domain (Cheyette et al., 1994; Kenyon et al., 2005). This amino acid change maps to the 4^{th} amino acid in the proposed decapeptide (figure 2.3.2.1.1). Kenyon *et al* show that this amino acid change "does not abolish DNA-protein interactions but alters co-factor binding specificity to resemble that of Optix" (Kenyon et al., 2005). Indeed whilst they show through gel shift experiments that SO^{D} protein can specifically shift a DNA fragment containing six tandem copies of the DNA target site for vertebrate Six1 and the requirement for a functional DNA-binding domain to induce the dominant mutant phenotype, they show through yeast-two hybrid interactions that

this amino acid change "alters the interaction profile of the Six domain and may lead to the abnormal recruitment of a putative partner of Optix" (Kenyon et al., 2005). Additionally, in the aim to show that it is the presence of a D residue and not the loss of a V that causes the SO^D phenotype, they show that mutating the 5^{th} amino acid in the decapeptide of Optix from a D to a G alters Optix protein function (figure 2.3.2.1.1) (Kenyon et al., 2005).

Another study of which the work supports in part the importance of the decapeptide I propose, is that of Kirby *et al* who in 2001 isolated a mutation of D-Six4, D-$Six4^{131}$, that is caused by the substitution of the 8^{th} amino acid of the decapeptide from an arginine to a cysteine (figure 2.3.2.1.1) (Kirby et al., 2001). D-$Six4^{131}$ mutants hatch but most die at larval and pupal stages (Kirby et al., 2001). Although, this mutation has only mild musculature defects, it results in a severe defect in gonadogenesis whereby SGPs and germ cells fail to coalesce (Kirby et al., 2001). Both the valine in the 4^{th} position of the decapeptide and the arginine in the 8^{th} position are conserved in all three subfamilies supporting their importance (figure 2.3.2.1.1).

Although more work is required to fully prove the importance of this decapeptide, the results obtained by these two different groups indicate that there is evidence for functional importance of this decapepetide.

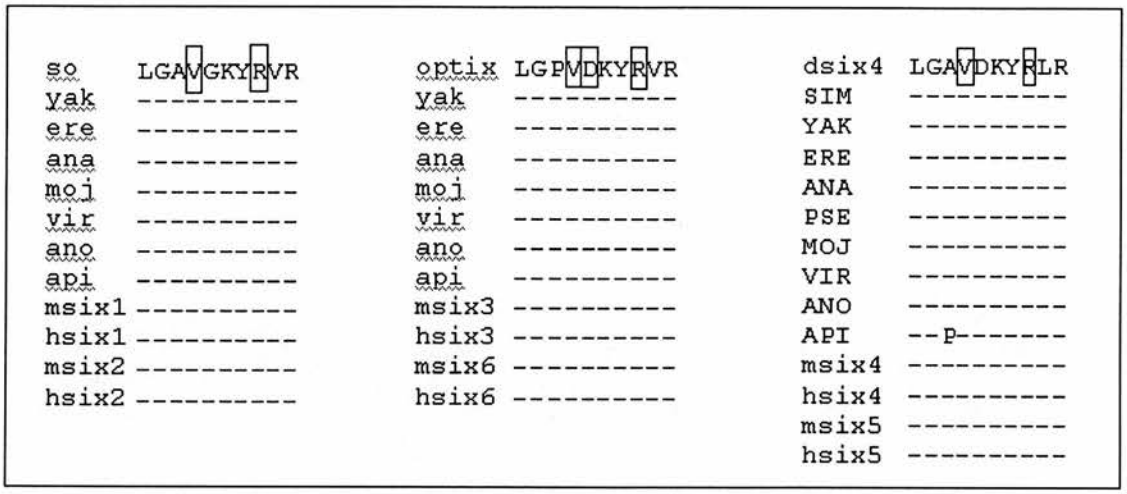


Figure 2.3.2.1.1: The Six domain decapeptide proposed for each of the three subfamilies. The amino acids for which studies have demonstrated functional importance are surrounded by a square box.

2.3.3 What are the additional criteria supporting the 1999 classification?

Previous sequence comparisons led to the classification into three subgroups of the Six family of homeodomain transcription factors (Seo et al., 1999). This classification was based on the conservation within the Six domain and the homeodomain of the proteins compared (Seo et al., 1999). Each subfamily was characterised according to a tetrapeptide found at the start of the homeodomain, adjacent upstream of the homeodomain characteristic tetrapeptide (Seo et al., 1999). The members of the Six3 family all share a very high level of sequence identity (>90%) and all display the tetrapeptide (QKTH) (Seo et al., 1999). Although their separation is less clear the Six2-like proteins are separated from the Six4-like according to the tetrapeptides (ETSY) and (ETVY) respectively (Seo et al., 1999). The members of the Six4 family have low characteristic substitutions in their homeodomains involving lysine and arginine at positions 17 and 24 respectively (Seo et al., 1999). Consistent with these criteria, I report that the characteristic tetrapeptides for each family are displayed consistently across all the sequences I compared in this study. I thereby support the 1999 classification.

In the light of the comparisons I carried out in this chapter, I established additional criteria in further support of this classification. The newly described Six domain characteristic decapeptide displays variability according to the family classification. I propose that the Six domain of the Six3/Six6 subfamily is characterised by the decapeptide (LG**P**V**D**KYR**V**R), the Six1/Six2 subfamily by (LG**A**V**G**KYR**V**R) and the Six4/Six5 subfamily by (LG**A**V**D**KYR**L**R).

I report that the *Apis mellifera* Six4 orthologue displays variation both in its homeodomain characteristic tetrapepetide (EVVY instead of ETVY) (Seo et al.,

1999) and in its Six domain characteristic decapeptide (LG**P**V**D**KYR**L**R instead of LG**A**V**D**KYR**L**R). The percentage of sequence identity of this domain with the *D. melanogaster* D-Six4 Six domain (68%) is the criterion on which I assigned this sequence to the Six4/Six5 subfamily. I therefore recommend that these subfamily specific decapeptides be used along side other criteria such as identity percentages for assigning Six domain sequences to specific subfamilies.

2.3.4 What is the relationship between the Six proteins in insects and mammals?

I showed that the percentages of sequence identities of the Six domains of the Six proteins of the insect and mammal species I compared and of the homeodomains of the Six proteins of the insect and mammal species I compared range from 56% to 96%.

"Insects can be subdivided into two evolutionary lineages, the Holometabola (insects with a complete metamorphosis during development) and Hemimetabola (insects with an incomplete metamorphosis)" (Hauser et al., 2006). *Apis mellifera* belongs to an insect order (the Hymenoptera), which has very recently been shown to occupy the most basal position in the Holometabola lineage (Savard et al., 2006). In contrast, the Diptera, of which the *Drosophila* species are a part of, are more recently evolved holometabolous insects (2006). From the results I obtained, I propose a functional conservation of these Six proteins in holometabolous insects and in the *Drosophila* genus in particular.

I showed that in mammals, the two paralogues of each of the three Six families are more similar to each other than to the corresponding *Drosophila*

homologues and that the two mammalian orthologues for each of the three Six families are identical both in the Six domains and the homeodomains (except for human and mouse Six5 which only show 99% identity in the Six domain) (tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3). It has been suggested that these six mammalian proteins have derived from gene duplications that occurred after the separation of deuterostomes from protostomes, possibly before or early in the vertebrate lineage (Seo et al., 1999). Nonetheless, I propose that the percentage of sequence identity between the mammal and insect proteins indicates potential functional redundancy.

2.3.5 What are the features of putative conservation?

I suggest that the percentages of identity between these sequences indicate a functional conservation of these Six proteins. The Six domain is thought to be involved both in determining DNA binding specificity and in mediating protein-protein interactions (Cheyette et al., 1994) (Chapter 1). The homeodomain is involved in protein-DNA interactions (Cheyette et al., 1994) (Chapter 1). I propose that varying levels of sequence conservation between the Six domains and the homeodomains indicate different levels of constraints with regards to the functioning of these domains. Overall, I show that the homeodomains of the Six proteins are more conserved than the Six domains. This could imply that the protein-protein and DNA-protein interactions mediated by the Six domain are more relaxed, allowing for interactions with different cofactors, than the protein-DNA interactions mediated by the homeodomain.

The DNA-binding sites of some Six proteins are known. Mouse Six5 and Six2 bind to the same sequence as does Six4, while Six3 does not (Kawakami et al.,

1996). Also, the putative DNA binding sequence of Six4 in *D. melanogaster* is identical to the SIX5 DNA binding sequence in humans (Kiosses, 2007, personal declaration). These results support the above suggestions.

In mammals, the percentages of sequence identity between paralogous homeodomains are much higher than between paralogous Six domains (tables 2.2.3.1-2.2.3.3 and 2.2.4.1-2.2.4.3). I suggest that the functional divergence of these paralogues is mainly due to the divergence in the Six domains and instead, the functions carried out by the homeodomains are more likely to have been conserved. The protein sequence of the Six domain of mouse Six4, shares a higher percentage of identity with the mouse Six1 paralogue (72%) than it does with the mouse Six5 paralogue (65%) (table 2.2.3.1). I suggest that the functional similarities between mouse Six4 and Six1 are probably related to their ability to interact with similar cofactors mediated by the Six domain as opposed to their ability to bind DNA, which would involve interactions with the homeodomain.

I propose that the sequence signatures characteristic of the Six domain and the homeodomain of the three subfamilies are functionally important.

2.3.6. What are the limitations of this study?

I have focused these sequence comparisons solely on the Six domains and the homeodomains of these proteins. Indeed, the N- and C-termini of these proteins are highly variable in length and show no sequence similarity both between the three *Drosophila* proteins (figure 2.3.6.1 and figure 2.3.6.2) and between the vertebrate homologues Six3 and Six4 (Seo et al., 1999). Consequently, I ignored these domains

in the investigation of the relationships between Six proteins of different animal groups.

The recently published results indicating that the replacement of the So C-terminal tail with the C-terminal tail of Optix has an inhibitory effect despite there being less than 10% amino acid similarity between these two regions suggests that these domains are likely to be critical in the functioning of these different proteins (Weasner et al., 2007). Further work requires to be carried out in order to assess their significance.

Also, more recent and very interesting results have been published showing the *in vitro* importance of the immediate downstream region of the homeodomains of both mouse Six2 and mouse Six6 in their binding affinities (Hu et al., 2008). Further work requires to be carried out in order to assess the *in vivo* significance of these results.

Figure 2.3.6.1: The N- and C- termini of the three *D. melanogaster* Six proteins, D-Six4, So and Optix vary considerably in length. The numbers indicate the beginning and end of the Six domain (SIX) and of the homeodomain (HD).

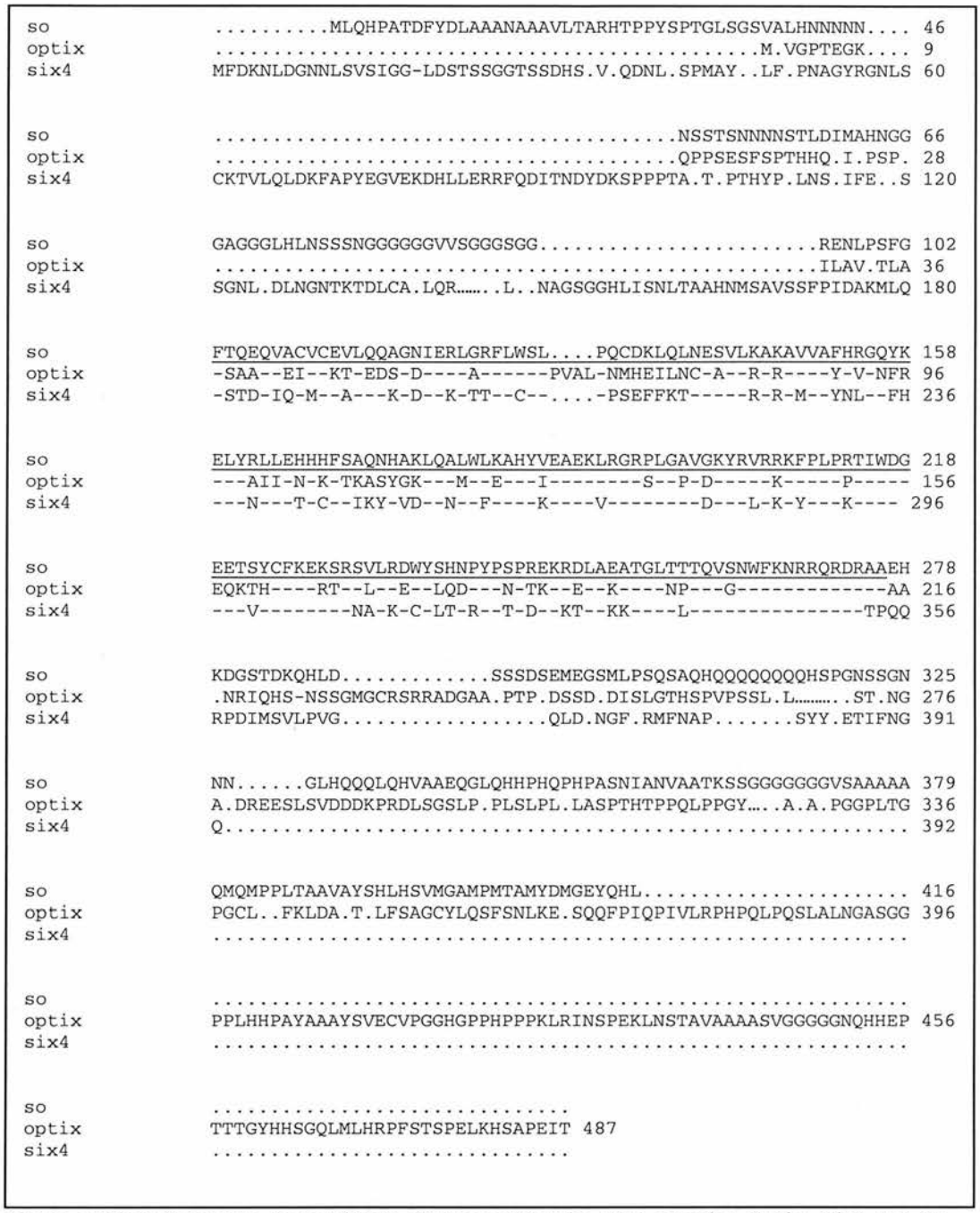


Figure 2.3.6.2: Sequence comparisons of *Drosophila* six class proteins Optix, Six4 and So. Sequence identities relative to So are indicated by hyphens; differing amino acids are shown. Gaps required for optimal alignments are represented by dots. The underlined sequence shows the start and the stop of the Six domain and the homeodomain. The N- and C- termini of these proteins show no similarity.

Chapter 3

The Six family phylogeny

The Six family phylogeny

3.1 Introduction

The evolution of animal diversity was accomplished largely by duplication and subsequent modification of existing genes and remodelling and redeployment of already existing genetic networks (Stierwald et al., 2004). Eye development and the expression of homologous members of the retinal determination gene network, of which the Six genes are a part, is an example of this (Silver and Rebay, 2005). Although the primitive eye of planaria, the compound eye of insects, and the camera-like eye of vertebrates are morphologically distinct, the molecular mechanisms that underlie their development are remarkably conserved (Silver and Rebay, 2005). It is therefore relevant to question how and when basic developmental networks were formed, only once for all phyla, or repeatedly when the evolutionary conditions were favourable.

Identifying orthologues in different organisms has become a lot easier since the genome sequencing revolution, which has resulted in the availability of a plethora of genomic sequences. In this chapter, I determined how ancestral the Six family of proteins is by collecting Six family protein sequences from basal metazoans. I applied the newly established criteria in order to define and classify these sequences. Further understanding of the Six family members relationship was also achieved.

3.1.1 How ancestral is the Six family of proteins?

Homologs of So from a diverse set of basal Metazoa, including representatives of the poriferan classes, the cnidarian classes and from platyhelminthes have previously been reported, indicating that So is highly conserved across the animal kingdom (Bebenek et al., 2004). Additionally, a detailed study of the Six class family has been

carried out in two species of jellyfish, *Cladonema radiatum*, and *Podocoryne carnea* (Stierwald et al., 2004). In both these species *Six* genes from the Six1/Six2 and Six3/Six6 subfamilies have been identified and their expression patterns analysed (Stierwald et al., 2004). A gene belonging to the Six4/Six5 subfamily has also been identified in *Cladomena radiatum* (Stierwald et al., 2004).

Four *Six* genes have previously been identified in *Caenorhabditis elegans*; Ceh 32, Ceh 33, Ceh 34 and Ceh 35 (Ruvkun and Hobert, 1998) (also known as SixC, SixB, SixA and SixD (Seo et al., 1999) and renamed NCESix3/6, NCESix1/2a, NCESix1/2b and NCESix4/5 in this study respectively). To date, very little is known about Ceh 33 and Ceh 34 but the expression patterns of Ceh 32 have been studied and preliminary functional studies of Ceh 35 have been carried out (Yanowitz et al., 2004). Importantly however, the nematode Six proteins seem to have diverged from the other members of the family (Seo et al., 1999).

In this chapter I report the identification of twenty-five Six protein sequences in six nematode species, four jellyfish species, two platyhelminth and one poriferan species. I compared these with the orthologuous Six proteins of *D. melanogaster* and characterised them accordingly.

3.1.2 How did I characterise the Six family proteins?

Previously, classification of *Six* genes into three different families was achieved by comparing percentages of sequence identity (Seo et al., 1999). In the previous chapter, I demonstrated that the Six family proteins are defined by a decapeptide in the Six domain with variations specific to each subfamily in addition to the previously described two adjacent tetrapetides in the homeodomain also with

subfamily specificities (Seo et al., 1999) (Chapter 2). I used these criteria for characterising and classifying the sequences obtained.

No comprehensive comparison of the Six family of protein sequences in basal metazoans has ever been carried out. Consequently, having characterised and classified the Six sequences collected, I was able to carry out sequence comparisons and yield further understanding of the relationships between these proteins.

3.2. Results

I collected and characterised the Six domain and homeodomain of the Six protein sequences for the different species and compared them with the *D. melanogaster* orthologues.

3.2.1 How did I collect the sequences?

I collected the basal metazoan sequences, in collaboration with Charles Clive, from the literature available and from various databases including The National Centre for Biotechnology Information (http://www.ncbi.nlm.nih.gov/), wormbase (http://www.wormbase.org/) and BaNG (http://www.nematodes.org/) (tables 3.2.1.1, 3.2.1.2 and 3.2.1.3).

3.2.2 How did I characterise the sequences?

Once collected, I classified the Six protein sequences into the three subfamilies using the criteria previously mentioned and further aligned and compared them as previously described (chapter 2) (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3). Neighbour-joining phylogenies with bootstraps were constructed using the Molecular Evolutionary Genetics Analysis software (MEGA) (Tamura et al., 2007).

Species	Gene ID	Also known as	Acronym in this thesis	Database used	Notes
Caenorhabditis elegans	Ceh32	SixC	NCESix3/6	Wormbase.org	Found with tblastn on BaNG
Caenorhabditis elegans	Ceh33	SixB	NCESix1/2a	Wormbase.org	Found with tblastn on BaNG
Caenorhabditis elegans	Ceh34	SixA	NCESix1/2b	Wormbase.org	Found with tblastn on BaNG
Caenorhabditis elegans	Unc39	SixD, Ceh35	NCESix4/5	Wormbase.org	Found with tblastn on BaNG
Caenorhabditis briggsae	CBG23282	CAE75312	NCBSix3/6	EBI SRS	Found with NCBI PSI-Blast, Ceh32 homologue
Caenorhabditis briggsae	CBG06764	CAE62634	NCBSix1/2a	EBI SRS	Found with NCBI PSI-Blast, Ceh33 homologue
Caenorhabditis briggsae	CBG06765	CAE62635	NCBSix1/2b	EBI SRS	Found with NCBI PSI-Blast, Ceh34 homologue
Caenorhabditis briggsae	CBG04628	CAE60911	NCBSix4/5	EBI SRS	Found with NCBI PSI-Blast, Unc39 homologue
Ascaris suum	BM033412	ASP07176_1	NASSix1/2	EBI SRS	Found with tblastn on BaNG
Ascaris suum	BM282217	ASP02982_1	NASSix3/6	EBI SRS	Found with annotation search on BaNG
Heterodena glycines	CB374918	HGP01951_1	NHGSix3/6	EBI SRS	Found with annotation search on BaNG
Meloidogyne chitwoodi	CB855601	MCP01070_1	NMCSix1/2	EBI SRS	Found with tblastn on BaNG
Pristionchus pacificus	AI986717	PPP00196_1	NPPSix3/6	EBI SRS	Found with tblastn on BaNG

Table 3.2.1.1: Information relating to the collection of the nematode sequences.

Species	Gene ID	Acronym in this thesis	Database used	Notes
Aurelia aurita	AY652604	JAASix1/2	EBI SRS	Missing very beginning of Six Domain
Cladonema radiatum	AY542529	JCRSix1/2	EBI SRS	(Stierwald et al., 2004)
Cladonema radiatum	AY542530	JCRSix3/6	EBI SRS	(Stierwald et al., 2004)
Cladonema radiatum	AY542531	JCRSix4/5	EBI SRS	(Stierwald et al., 2004)
Podocoryne carnea	AY542527	JPCSix1/2	EBI SRS	(Stierwald et al., 2004)
Podocoryne carnea	AY542528	JPCSix3/6	EBI SRS	(Stierwald et al., 2004)
Hydra magnipapillata	CV464195	JHMSix3/6a	EBI SRS nucleotide	Translated from nucleotide sequence
Hydra magnipapillata	DN812339	JHMSix3/6b	EBI SRS nucleotide	Translated from nucleotide sequence

Table 3.2.1.2: Information relating to the collection of the jellyfish sequences.

Species	Gene ID	Acronym in this thesis	Database used	Notes
Dugesia japonica	CAD89530	PDJSix1/2	EBI SRS	Found with NCBI annotation search
Girardia tigrina	CAB89515	PGTSix1/2	EBI SRS	Found with NCBI annotation search
Girardia tigrina	AAN77127	PGTSix3/6	EBI SRS	Found with NCBI annotation search
Haliclona	AAT69264	SHSix1/2	EBI SRS	Found with NCBI annotation search

Table 3.2.1.3: Information relating to the collection of the platyhelminth and sponge (*Haliclona)* sequences.

Acronym in this thesis	% identity with So SD/HD	% identity with Optix SD/HD	% identity with Six4 SD/HD	Assigned subclass
NCESix3/6	47/67	54/88	53/59	Six3/6
NCESix1/2a	57/80	47/60	48/68	Six1/2
NCESix1/2b	52/67	40/52	50/59	Six1/2
NCESix4/5	36/50	38/45	37/52	Six4/5
NCBSix3/6	48/67	54/88	51/59	Six3/6
NCBSix1/2a	56/80	47/61	50/68	Six1/2
NCBSix1/2b	52/68	41/50	52/65	Six1/2
NCBSix4/5	38/37	34/32	41/37	Six4/5
NASSix1/2	67/NA	55/NA	55/NA	Six1/2
NASSix3/6	NA/65	NA/88	NA/60	Six3/6
NHGSix3/6	32/NA	40/NA	33/NA	Six3/6
NMCSix1/2	50/NA	40/NA	40/NA	Six1/2
NPPSix3/6	42/NA	46/NA	38/NA	Six3/6

Table 3.2.2.1: Percentages of identity of the nematode sequences obtained and their classification in the most probable subfamily. NA stands for non applicable signifying that no sequences were obtained for those domains.

Acronym in this thesis	% identity with So SD/HD	% identity with Optix SD/HD	% identity with Six4 SD/HD	Assigned subclass
JAASix1/2	76/85	69/62	63/72	Six1/2
JCRSix1/2	69/85	60/63	55/70	Six1/2
JCRSix3/6	58/68	63/84	60/57	Six3/6
JCRSix4/5	59/68	50/59	54/68	Six4/5
JPCSix1/2	72/86	63/63	54/70	Six1/2
JPCSix3/6	61/68	66/85	56/57	Six3/6
JHMSix3/6a	58/52	62/76	54/47	Six3/6
JHMSix3/6b	58/53	62/78	54/46	Six3/6

Table 3.2.2.2: Percentages of identity of the jellyfish sequences obtained and their classification in the most probable subfamily.

Acronym in this thesis	% identity with So SD/HD	% identity with Optix SD/HD	% identity with Six4 SD/HD	Assigned subclass
PDJSix1/2	72/90	56/63	54/72	Six1/2
PGTSix1/2	72/88	56/62	54/70	Six1/2
PGTSix3/6	60/67	68/88	56/62	Six3/6
SHSix1/2	67/88	59/63	54/70	Six1/2

Table 3.2.2.3: Percentages of identity of the platyhelminth and sponge sequences obtained and their classification in the most probable subfamily.

3.2.3. What did the comparisons of the Six domains show?

I aligned the Six domains of the Six proteins collected (figures 3.2.3.1, 3.2.3.2 and 3.2.3.3). I showed that the Six domain of the only Six protein found in *Haliclona* displays 67% identity to So Six domain (table 3.2.2.3). I report that orthologues of So show the highest percentages of identity with So (compared to the percentages of identity among the orthologues of D-Six4 and Optix) among nematodes, jellyfish, and platyhelminthes (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3). I show that the percentages of identity of these range from 50% to 67% for the nematode species, from 69% to 76% for the jellyfish species and 72% for the platyhelminthes species (tables 3.2.2.1 and 3.2.2.3). I report that orthologues of D-Six4 show the lowest percentages of identity with D-Six4 (compared to the percentages of identity between the orthologues of So and Optix). I show that the percentages are 37% and 41% in the nematode species and 54% in the jellyfish species (tables 3.2.2.1 and 3.2.2.2). Overall, I report that the percentages of identity between the orthologous Six domains are highest in the jellyfish and platyhelminthes (tables 3.2.2.2 and 3.2.2.3) and that the Six domains found in nematode species show a lower percentage of identity (3.2.2.1).

I derived phylogenetic trees from these sequence comparisons (figures 3.2.3.4 and 3.2.3.5). The phylogenetic tree for the Six3/Six6 subfamily did not provide meaningful results. The tree obtained did not display any evolutionary time between the evolution of the protein sequences of the jellyfish and those of humans. These problems are a default of the software resulting from difficulties in dealing with high levels of sequence identity and consequently this tree was excluded from further analysis. Previous studies of *C. elegans* Six proteins show that they exhibit lower

percentages of identity relative to the other Six proteins (Seo et al., 1999). The homeodomains in particular are quite divergent and show single amino acid substitutions in the tetrapetides relative to those from other species (Seo et al., 1999). I therefore used the nematode species as an outgroup to root the trees (figures 3.2.3.4 and 3.2.3.5). Although the position of nematode sequences in the trees looks odd, the use of the nematode species as an outgroup was verified and supported by Deborah Charlesworth (personal communication).

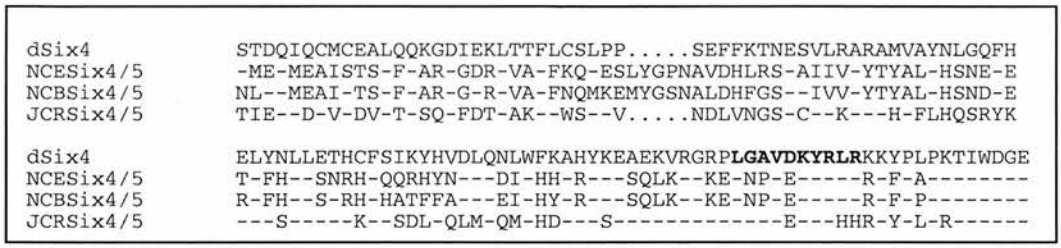


Figure 3.2.3.1: Sequence alignment of the Six domain of *D. melanogaster* D-Six4 and its orthologues in nematode and jellyfish species. In bold is the decapeptide characteristic of the Six4/Six5 family. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

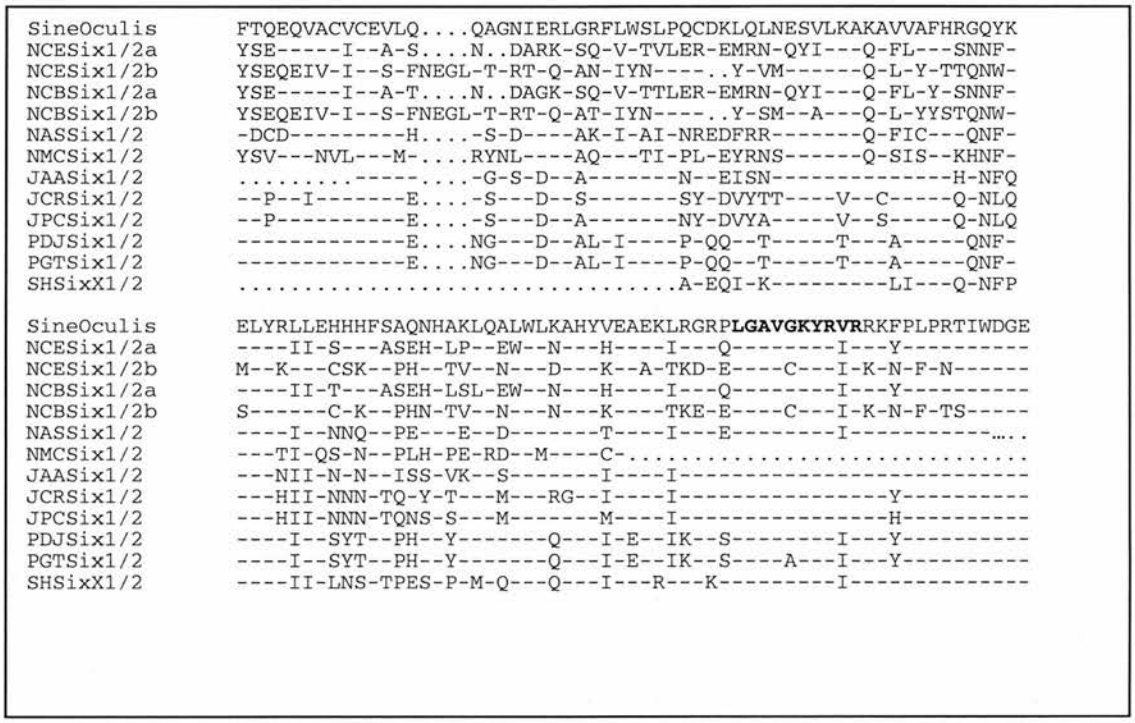


Figure 3.2.3.2: Sequence alignment of the Six domain of *D. melanogaster* So and its orthologues in nematode, jellyfish, platyhelminth species and *Haliclona.* In bold is the decapeptide characteristic of the Six1/Six2 family. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

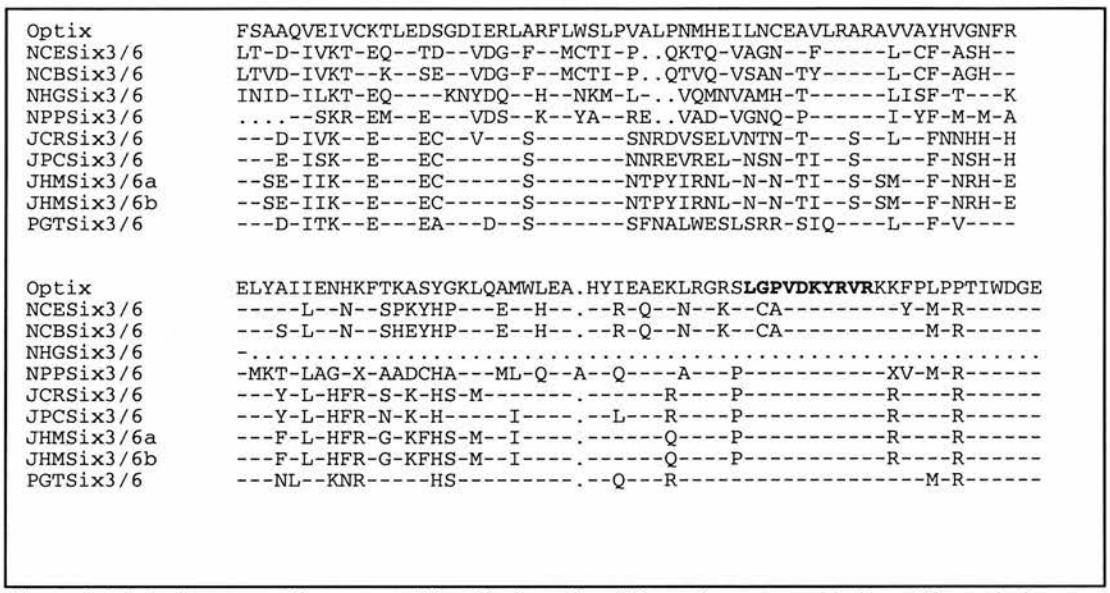


Figure 3.2.3.3: Sequence alignment of the Six domain of *D. melanogaster* Optix and its orthologues in nematode, jellyfish, platyhelminth species. In bold is the decapeptide characteristic of the Six3/Six6 family. Sequence identities relative to Optix are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

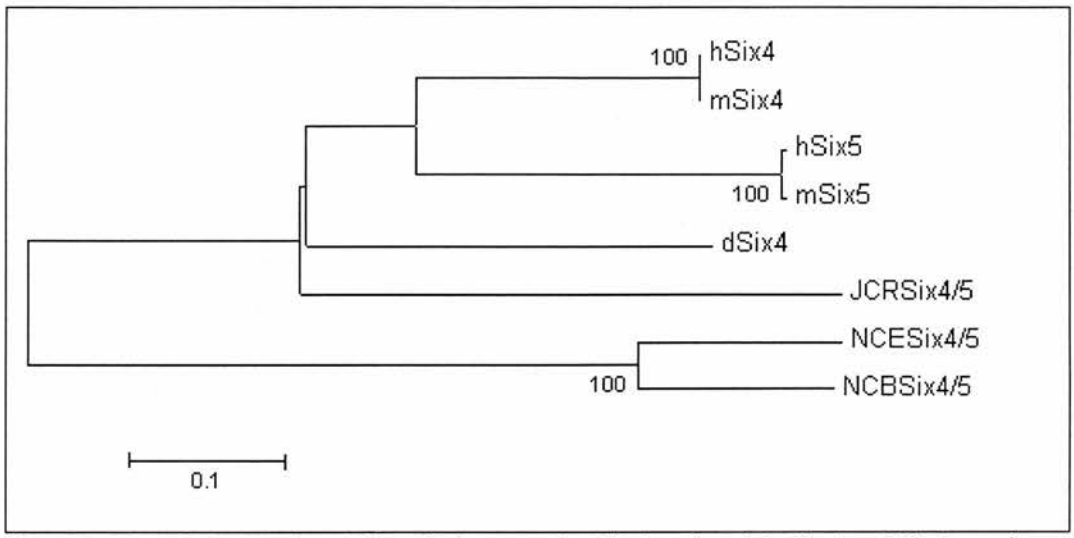


Figure 3.2.3.4: Phylogenetic relationship between the Six domain of the Six4 and Six5 proteins of jellyfish and nematode species, *Drosophila melanogaster* and mammals.

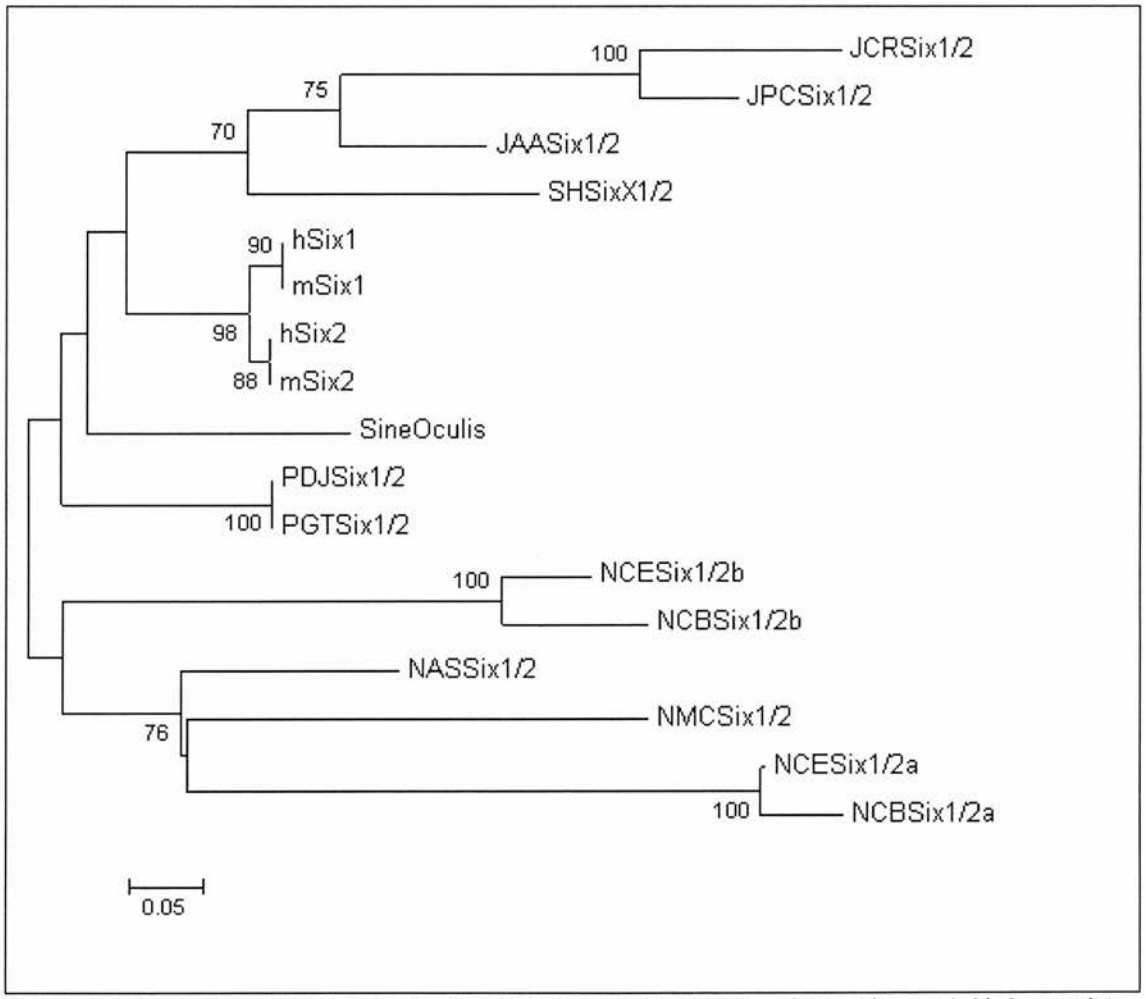


Figure 3.2.3.5: Phylogenetic relationship between the Six domain of the Six1 and Six2 proteins jellyfish, nematode and platyhelminth species, *Haliclona, Drosophila melanogaster* and mammals.

3.2.4 What did the comparisons of the homeodomains show?

I aligned the homeodomains of the Six proteins collected (figures 3.2.4.1, 3.2.4.2 and 3.2.4.3). I showed that the homeodomain of the only Six protein found in *Haliclona* displays 88% identity to So (table 3.2.2.3). I report that the percentages of identity are much higher between the homeodomains of the three *Drosophila melanogaster* proteins and their orthologues in these species, than they were for the Six domains (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3). I showed that they range from 37% to 100% in nematodes, 68% to 86% in jellyfish and 88% to 90% in platyhelminthes (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3). Similarly to the Six domain, I report that the percentages of identity are the lowest in the orthologues of D-Six4 (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3). I derived phylogenetic trees from these results (figures 3.2.4.4 and 3.2.4.5). No tree was derived for the Six1/Six2 subfamily. Again I used the nematode sequences as an outgroup. Once again, the use of the nematode species as an outgroup was supported by Deborah Charlesworth (personal communication).

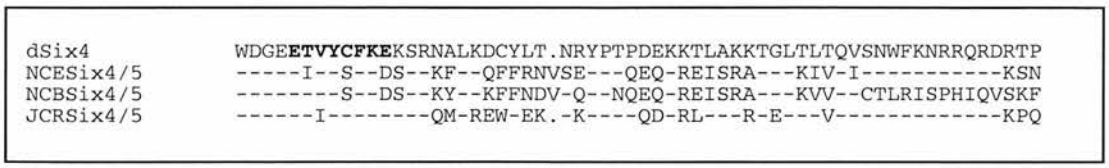


Figure 3.2.4.1: Sequence alignment of the homeodomain of *D. melanogaster* D-Six4 and its orthologues in nematode and jellyfish species. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots. In bold are the tetrapeptide characteristic of the Six4/Six5 family adjacent to the tetrapeptide characteristic of the Six class homeodomains respectively.

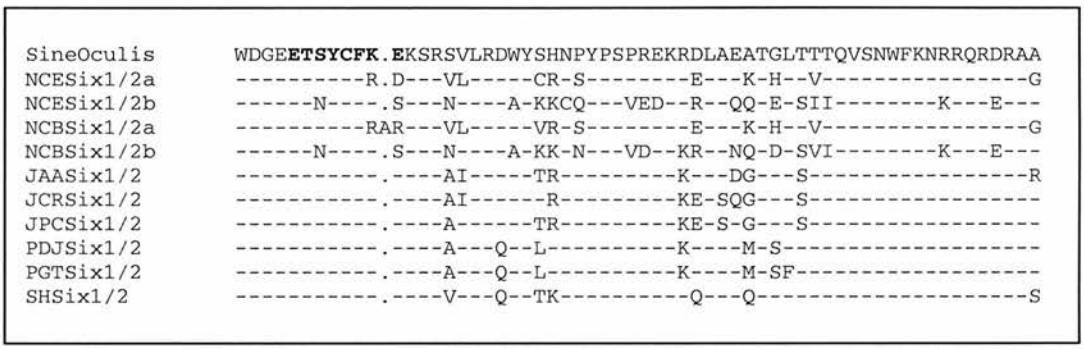


Figure 3.2.4.2: Sequence alignment of the homeodomain of *D. melanogaster* So and its orthologues in nematode, jellyfish, platyhelminth species and *Haliclona*. Sequence identities relative to So are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots. In bold are the tetrapeptide characteristic of the Six1/Six2 family adjacent to the tetrapeptide characteristic of the Six class homeodomains respectively.

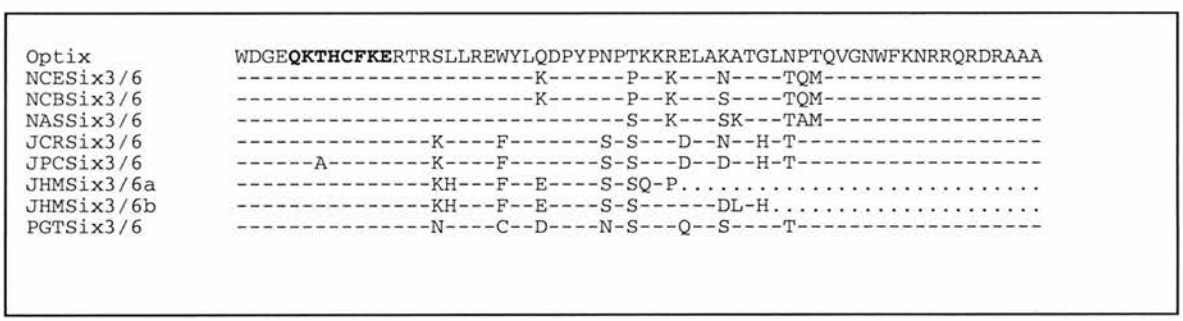


Figure 3.2.4.3: Sequence alignment of the homeodomain of *D. melanogaster* Optix and its orthologues in nematode, jellyfish, platyhelminth species. Sequence identities relative to Optix are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots. In bold are the tetrapeptide characteristic of the Six3/Six6 family adjacent to the tetrapeptide characteristic of the Six class homeodomains respectively.

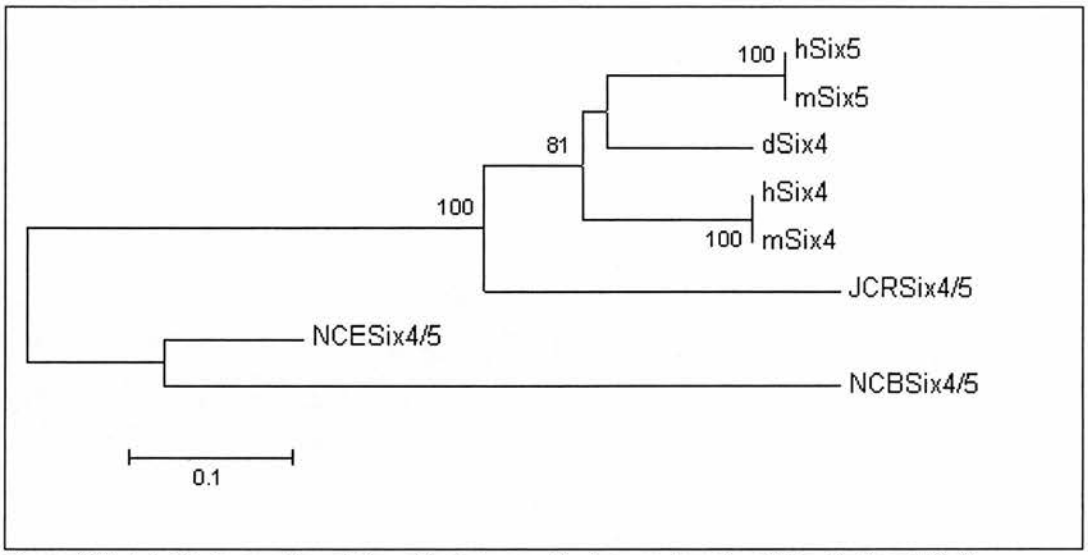


Figure 3.2.4.4: Phylogenetic relationship between the homeodomain of the Six4 and Six5 proteins of nematode and jellyfish species, *Drosophila melanogaster* and mammals.

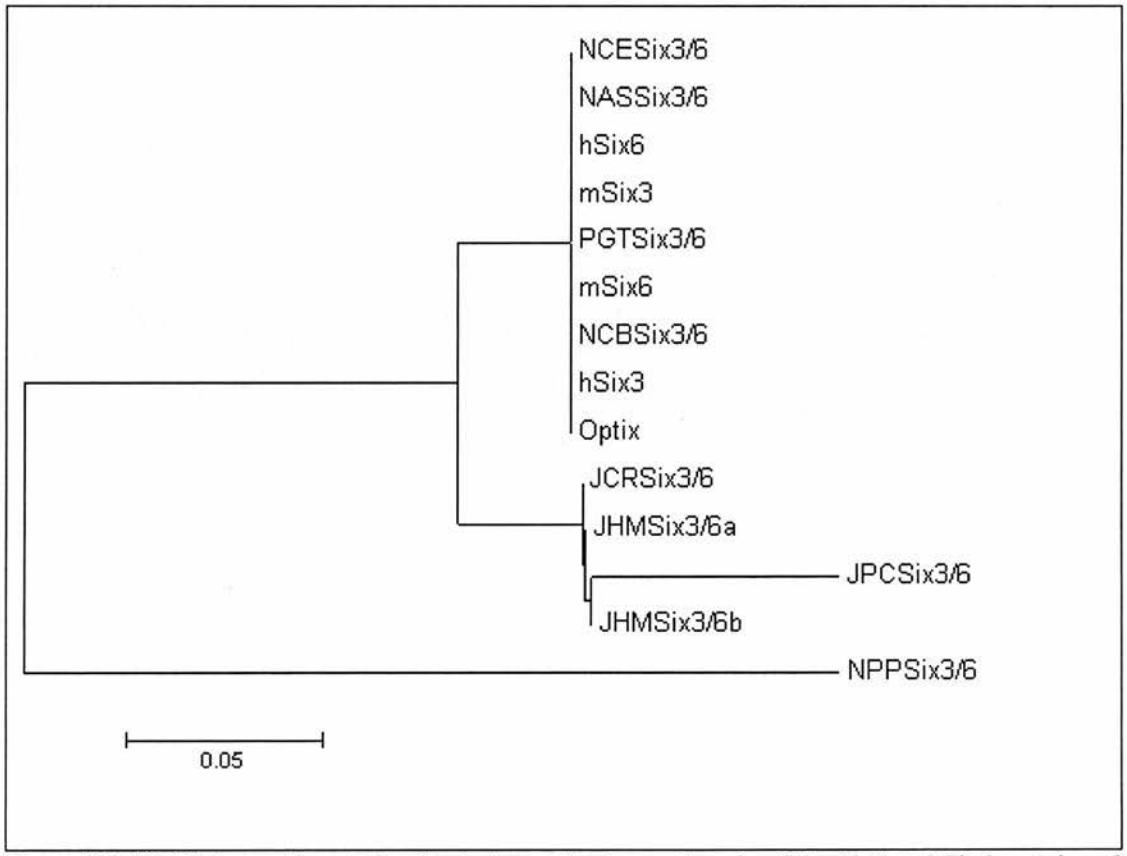


Figure 3.2.4.5: Phylogenetic relationship between the homeodomain of the Six3 and Six6 proteins of nematode, jellyfish and platyhelminth species, *Drosophila melanogaster* and mammals.

3.3 Discussion

In this chapter, I identified the Six domains and homeodomains of Six protein sequences of six nematode, four jellyfish, two platyhelminth and one poriferan species. I thereby established the conservation of the Six family of proteins from poriferans to mammals. I characterised these proteins in these basal metazoans and thus further understanding of the relationship between the different members of the Six family of proteins was achieved. I discuss this below.

3.3.1 How ancestral is the Six family of proteins?

Three *Six* genes and six *Six* genes have previously been identified in Insects and mammals respectively (Chapter 1). The identification and characterisation of Six protein sequences in nematode, jellyfish, platyhelminth and poriferan species is critical in establishing the ancestry of these proteins. I only identified one Six protein in *Haliclona,* a member of the Six1/Six2 family. This could be an artefact of the present lack of data. Alternatively the Six proteins belonging to the Six1/Six2 subfamily are the earliest Six proteins to have evolved. Indeed, poriferans are among the earliest metazoans to have evolved. I identified no orthologue of the Six4/Six5 family in the species of platyhelminth. Again, this could reflect the current lack of data. Alternatively, this may suggest a later evolution of this subfamily. I identified an orthologue of the Six4/Six5 family in one jellyfish species. Currently, jellyfish are the earliest species in which a protein from the Six4/Six5 subfamily has been identified. Consistent with previous results, I report that the nematode sequences that I collected and compared are more distantly related to the other Six protein members indicating that they have diverged from the rest of the Six family (Seo et al., 1999).

As a result of the collection and analysis of this data, I have established the conservation of the Six family of proteins from poriferans to mammals.

3.3.2 How did I characterise the Six family?

Consistent with previous comparisons, the homeodomains of the Six protein sequences I compared show higher percentages of identity and less variability than do the Six domains (tables 3.2.2.1, 3.2.2.2 and 3.2.2.3) (Seo et al., 1999) (Chapter 2). Nonetheless, I report that the Six domain characteristic decapeptide is overall conserved (figures 3.2.3.1, 3.2.3.2 and 3.2.3.3). Nematode sequences aside, I showed that the Six3/Six6 decapeptide is conserved in the jellyfish and platyhelminthes species I compared (figure 3.2.3.3) and the Six1/Six2 decapeptide displays only up to two substitutions in these latter two species and only one substitution in *Haliclona* (figure 3.2.3.2). I showed that the Six4/Six5 decapeptide is poorly conserved (figure 3.2.3.1).

Nematode sequences aside, I report that the previously described homeodomain unique peptide sequence (CFKE) adjacent to the family specific tetrapeptide is displayed in all sequences I collected (figures 3.2.4.1, 3.2.4.2 and 3.2.4.3)(Seo et al., 1999). I report that the previously described characteristic tetrapeptides for each family are also displayed consistently across most of the sequences I compared (figures 3.2.4.1, 3.2.4.2 and 3.2.4.3) (Seo et al., 1999). I showed that two jellyfish species, *Cladonema radiatum* and *Podocoryne carnea* display single amino acid substitutions in their tetrapetides relative to the Six4/Six5 and Six3/Six6 families respectively (ETIY) and (QKAH) (figures 3.2.4.1, 3.2.4.2 and 3.2.4.3). Additionally, the twelve identical amino acids, that I showed were identical

across the four consensus sequences are identical in all of the sequences I compared in the Six3/Six6 and Six4/Six5 subfamily (figures 3.2.4.1 and 3.2.4.3) (Chapter 2). I report that the Six1/Six2 subfamily overall shows identity with ten of these twelve amino acids (figure 3.2.4.2). I consequently have established the conservation of these sequence signatures from poriferans to mammals.

Consistent with previous comparisons (Seo et al., 1999) the nematode sequences I compared in this study do not match these criteria and are divergent with regards to the other Six proteins sequences. The percentage of sequence identity of SixC and SixB has been previously shown to be lower than in the other Six proteins (Seo et al., 1999). SixB homeodomain in particular is quite divergent relative to other members of the Six2 family(Seo et al., 1999). SixA and SixD have homeodomains with single amino acid substitutions in their tetrapetides relative to the Six2 and Six4 families respectively, (ETNY) and (EIVY) (Seo et al., 1999). These two homeodomains also have approximately the same sequence identity relative to both families (Seo et al., 1999). Hence on the basis of the homeodomains sequences alone, the nematode SixA and SixD proteins cannot be clearly assigned to any of the three major families of the Six class (Seo et al., 1999). Whereas SixC belongs to the Six3 family, the Six domains of SixA, SixB and SixD exhibit similarities to both the Six2 and Six4 families (Seo et al., 1999).

I report that the other nematode species too show lower percentages of sequence identity relative to other Six proteins (table 3.2.2.1). I showed that the Six domain Six1/Six2 and Six3/Six6 decapeptides both show two substitutions conserved in both *C. elegans* and *C. briggsae* (LGAV**C**KYR**I**R) and (L**CA**VDKYRVR) respectively. The Six domain Six4/Six5 decapeptide shows three substitutions also

conserved in both *C. elegans* and *C. briggsae* (LNPVEKYRLR). Additionally, I report that the Six1/Six2 decapeptide in *Ascaris suum* shows one of the substitutions conserved in *C elegans* and *C. briggsae* (LGAVGKYRIR). I report however that the *Ascaris suum* Six3/Six6 decapeptide, identified solely by its location in the alignments, is completely different (.GPL.IISCS).

I report that the homeodomain tetrapeptides specific to each subfamily are conserved in the nematode species except for the *C.elegans* ones discussed and for the Six1/Six2 specific tetrapeptide in *C. briggsae* which I report have the same substitution as *C. elegans* (ETNY). I report that the homeodomain unique peptide sequence (CFKE) adjacent to the family specific tetrapeptide is not as well conserved and shows substitutions in up to three positions. These results clearly suggest divergence of these proteins in nematode species. Whether these sequence divergences have resulted in functional differences will require future work.

3.3.3 What is the relationship between the Six family members?

Little is known about the functions of the Six proteins in basal metazoans. Evidence of sensory structures, presence of a So homologue in sponges and aspects of expression of So homologues in some basal Metazoa suggest a functional role in the development of sense organs in these taxa(Bebenek et al., 2004). The *Girardia tigrina so* gene is expressed in the fully developed eye suggesting a putative role in development and maintenance of the eye(Pineda et al., 2000). *so* transcripts have been observed in the rhopalia sensory structures of *Aurelia medusa(Bebenek et al., 2004)*. Rhopalia have been shown to be important in orientation and sensing gravity and aid in recovery of tilting during swimming (Bebenek et al., 2004). Expression in

the rhopalia is consistent with a possible sensory role for So during the development of the eye (Bebenek et al., 2004). In adult *Aurelia*, *so* transcripts have also been observed in the manubrium, the region around the mouth bearing the oral arms that is likely to have tactile and olfactory sensory functions (Bebenek et al., 2004). Additionally, Six1/Six2 and Six3/Six6 homologues are expressed in the eye cup of *Cladonema radiatum*, a jelly fish species with well-developed lens eyes in the tentacle bulbs (Stierwald et al., 2004) and in *Podocoryne carnea*, a jellyfish without eyes, Six1/Six2 and Six3/Six6 are also expressed in the tentacle bulbs (Stierwald et al., 2004). I suggest that these findings may indicate a conserved role for So and Optix homologues from poriferans to mammals in the development of sense organs (Chapter 1).

In *Cladonema radiatum* Six4/Six5 homologue is mainly expressed in the manubrium, the feeding, and sex organ but is absent in the eye cup (Stierwald et al., 2004). Together with Six4/Six5 homologue, Six1/Six2 and Six3/Six6 homologues are expressed in the striated muscle (Stierwald et al., 2004). Six1/Six2 gene appears to be involved in both the myogenic/mesodermal (striated muscle) and the neurogenic line (nerve, eye) (Stierwald et al., 2004). In *Podocoryne carnea*, Six1/Six2 is also expressed in the manubrium and striated muscle, and Six3/Six6 in the mechanosensory nematocytes of the tentacle (Stierwald et al., 2004). In *Aurelia,* the oral arms where *so* transcripts have been observed also function in reproduction, storing the gametes and planula larvae and coordinating their release (Bebenek et al., 2004). I suggest that these findings may indicate a conserved role for homologues of the Six4/Six5 subfamily in the development of mesodermal structures and an

ancestral role for homologues of the Six1/Six2 and Six3/Six6 subfamilies in these taxa.

In *C. elegans,* the expression of *ceh-32* begins during embryogenesis in the hypodermal and neuronal precursor cells of the head (Dozier et al., 2001). Upon hatching, *ceh-32* is expressed in head hypodermal cells and neurons as well as in gonadal sheath cells (Dozier et al., 2001). RNAi inactivation studies suggest that *ceh-32* plays a role in head morphogenesis. This is consistent with expression patterns of Optix in *Drosophila melanogaster* and I suggest a conservation of function of Optix homologues in nematodes and insects.

Mutations in the *unc-39* gene (also known as Ceh 35) of *C. elegans* lead to migration and differentiation defects in a subset of mesodermal and ectodermal cells, including muscles and neurons (Yanowitz et al., 2004). Defects range from the disruption of mesodermal specification and differentiation to a disruption in neuronal migration and axon pathfinding (Yanowitz et al., 2004). This is consistent with expression patterns of D-Six4 in *D. melanogaster* and suggests a conservation of function of Six4 homologues in nematodes and insects. Additionally, the human Six5 Six domain and homeodomain can complement mutations in unc-39 (Yanowitz et al., 2004). This indicates that human Six5 and *C. elegans* UNC-39 are functional homologs (Yanowitz et al., 2004). These results indicate that some of the DNA binding specificity and some of the protein–protein interactions must have been maintained during the evolution of nematodes and humans (Yanowitz et al., 2004).

I suggest that despite the sequence divergence of the nematode Six proteins these findings indicate that there are aspects of functional conservation between these

Six proteins. A thorough functional comparison between Six proteins in nematodes and insects will shed light on these relationships.

Chapter 4

Features of putative functional importance: specificities of the three subfamilies

Features of putative functional importance: specificities of the three subfamilies

4.1 Introduction

In previous chapters I used protein sequence comparisons to define the Six family of proteins. I focused the comparisons on the similarities between orthologues. I used the percentage of identities and the conservation of sequence signatures as means of identifying Six proteins and I established their phylogeny. I proposed that these sequence signatures had a functional role. Additionally, I used these signatures to assign the Six family proteins into three subfamilies. In this chapter, I focused the comparisons on the differences between the three subfamilies in an aim to highlight additional features of putative functional importance.

4.2 Results

I aligned the Six domain and homeodomain of the three *Drosophila* paralogues and I identified amino acids specific to each subfamily. I then assessed the conservation of these amino acids in the species I had compared in the two previous chapters in order to establish their putative functional importance.

4.2.1 How did I identify amino acids specific to each subfamily?

I classed amino acids into six groups according to conserved properties (table 4.2.1.1). I used these groups to identify fourteen amino acids whose properties were not conserved and that were different in all three proteins. I assigned these as representatives of the specificity of each subfamily (figure 4.2.1.1).

Property	Amino acids in this group
Aliphatic	I, L, V
Aromatic	F, Y, W, H
Acidic	D, E
Basic	R, K
Tiny	A, C, G, T, S
Non of the above	Q, P, N, M

Table 4.2.1.1: I grouped the amino acids in six groups according to conserved properties. These are detailed above.

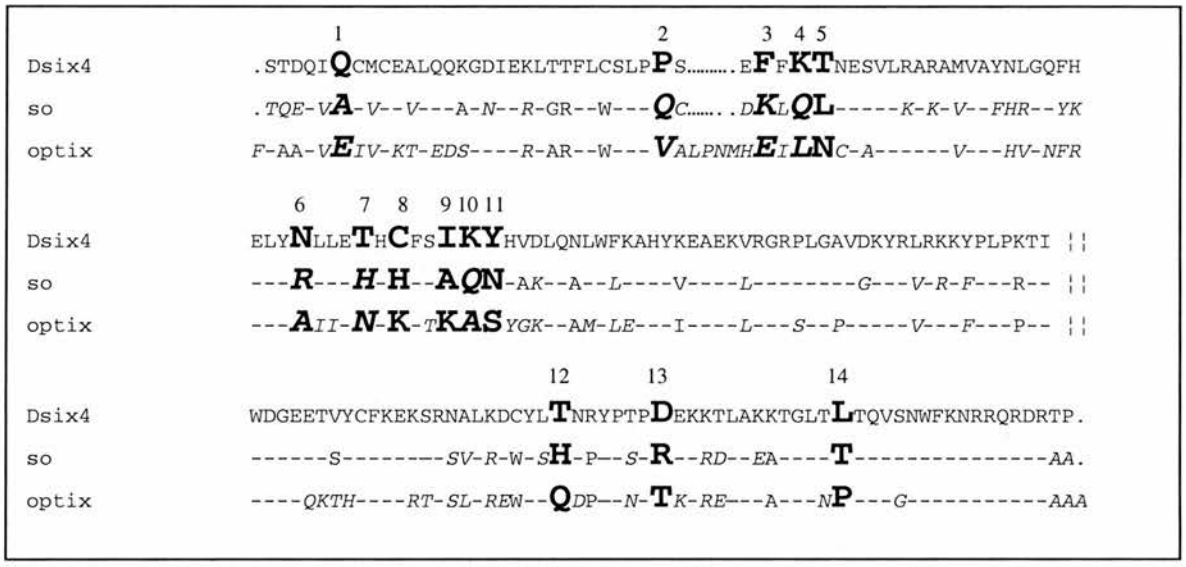


Figure 4.2.1.1: Sequence alignment of the Six domain and homeodomain of So, Optix and D-Six4. The amino acids in bold are the fourteen amino acids selected for further comparisons. The vertical dashed lines indicate the end of the Six domain and start of the homeodomain. Sequence identities relative to D-Six4 are indicated by hyphens and the differing amino acids are shown. Gaps required for optimal alignments are represented by dots.

4.2.2 How conserved are these amino acids unique to each subfamily?

I tabulated the substitutions of these fourteen unique amino acids in the different species I had previously compared (tables 4.2.2.1, 4.2.2.2 and 4.2.2.3).

I report that seven of the selected amino acids remain identical to the one present in *D. melanogaster* in over half the number of species compared in the case of D-Six4 and of So and eight in the case of Optix (figure 4.2.2.1). Additionally, I show that when the amino acids have been substituted, their properties as grouped in table 4.2.1.1 have been conserved.

I show that Leucine in position 336 in D-Six4, Arginine in position 245 in So and proline in position 197 in Optix remain identical throughout Metazoa except in two species of nematodes, *C.elegans* and *C. briggsae,* where they each differ (tables 4.2.2.1, 4.2.2.2 and 4.2.2.3). I show that Leucine in position 336 in D-Six4 and Lysine in position 106 in Optix when different to the amino acid in *D. melanogaster* are substituted consistently for aliphatic amino acids in the case of D-Six4 orthologues and basic amino acids in the case of Optix orthologues, thus conserving the property exhibited in *D. melanogaster* (tables 4.2.2.1 and 4.2.2.3). I show that the non aliphatic amino acid Tyrosine in position 258 in So is present in all species except in four species of nematodes where it is consistently substituted for an aliphatic amino acid (table 4.2.2.2). I show that Tyrosine in position 244 in D-Six4 and Tyrosine in position 184 in Optix when different to the amino acid in *D. melanogaster* are almost consistently substituted for a Serine (tables 4.2.2.1 and 4.2.2.3).

The amino acids in the three tables align when the three protein sequences are compared (figure 4.2.1.1). They are thus in the same position respectively in all

three proteins. I report that the 14th amino acid that I selected is the amino acid that remains identical in the most species to the one present in *D. melanogaster* overall (tables 4.2.2.1, 4.2.2.2 and 4.2.2.3). I also report that the first amino acid that I selected remains identical to the one shown in So in *D. melanogaster* in all but three orthologous proteins and in Optix and D-Six4 orthologues when it is substituted it is often replaced by an amino acid of the same property as shown in So (tables 4.2.2.1-4.2.2.3).

I also show that amino acids positioned in the homeodomain remain identical in more species than those positioned in the Six domain (figure 4.2.2.1). Finally, I show that, according to the criteria that I set, there are almost twice as many amino acids in the Six domain that are different in all three proteins, D-Six4, So and Optix, and that show no conserved properties (11/116*100= 9.5%) than there are in the homeodomain (3/60*100= 5%).

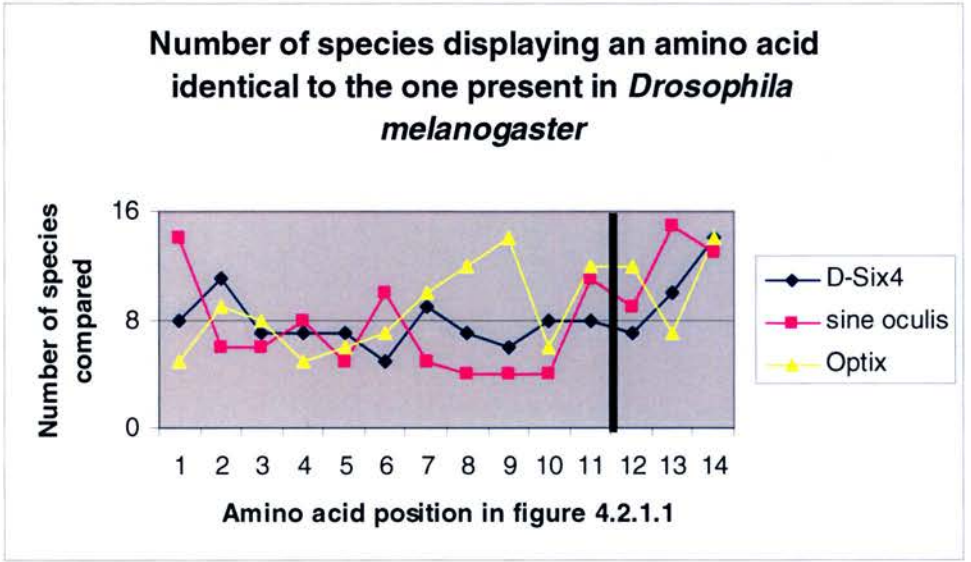


Figure 4.2.2.1: Number of species showing identical amino acids to *D. melanogaster* D-Six4, So and Optix. The vertical line indicates where the Six domain ends and where the homeodomain begins. The number of species compared was greater both for the Six1/Six2 and Six3/Six6 subfamily. The numbers were standardised to the lowest number of species compared, 16 in the Six4/Six5 subfamily.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
DSix4	**Q^{187}**	**P^{211}**	**F^{194}**	**K^{216}**	**T^{217}**	**N^{240}**	**T^{244}**	**C^{246}**	**I^{249}**	**K^{250}**	**Y^{251}**	**T^{316}**	**D^{323}**	**L^{336}**
D.simulans														
D.yakuba														
D.erecta														
D.ananassae														
D.pseudoobscura														
D.mojavensis						S								
D.virilis						S			M					
Anopheles gambiae			L	S	G	A	S	F	P			R		
Apis mellifera	S		L	R	R	S	S	P	S	D	R	R	E	
Mouse Six4	A	Q	L	R	G	S	S	S	S	A	N	Q	A	
Human Six4	A	Q	L	R	G	S	S	S	S	A	N	Q	A	
Mouse Six5	A		R	R	G	Q	S	P	A	A	H	G		
Human Six5	A		R	R	G	R	S	P	A	A	H	G		
Nematodes C.elegans	*E*	*S*	*D*	*L*	*R*	*H*	*N*	*H*	*Q*	*R*	*H*	*N*	*E*	*I*
Nematodes C.briggsae	*E*	*E*	*D*	*F*	*G*	*H*		*H*	*A*	*T*	*F*	*D*	*E*	*V*
Jellyfish C.radiatum	*D*	*V*	*L*	*N*	*G*	*H*		*K*	*S*	*D*	*L*	*K*	*Q*	

Table 4.2.2.1: Conservation through evolutionary time of key amino acids of *D. melanogaster* Six domain and homeodomain of D-Six4 protein. The first row refers to the numbers in figure 4.2.1.1. The second row shows the amino acids that were in bold in figure 4.2.1.1 in D-Six4. Their respective numbers correspond to their position with regards to the whole D-Six4 protein. The following rows show the amino acids at those positions in all the species previously used in the sequence comparisons. The empty cells indicate that the amino acid is identical to the one in *D. melanogaster* D-Six4 sequence. The letters in the other cells show which amino acid is present. The colours have been added to indicate what characteristic across time has been conserved in the amino acid substitutions (red for aromatic, pink for aliphatic, green for tiny, blue for acidic, dark blue for basic and no colour for other). The vertical line indicates where the Six domain ends and where the homeodomain begins.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Sine oculis	A^{108}	Q^{132}	K^{115}	Q^{137}	L^{138}	R^{161}	H^{165}	H^{167}	A^{170}	Q^{171}	N^{172}	H^{238}	R^{245}	T^{258}
D.yakuba														
D.erecta														
D.ananassae														
D.pseudoobscura														
D.mojavensis														
D.virilis														
Anopheles gambiae								Q	P	H				
Apis mellifera		A	R	H	R		S	T	P	H		T		
Mouse Six1		A	H	H	K	K	S	Q	P	H				
Human Six1		A	H	H	K	K	S	Q	P	H				
Mouse Six2		A	H	H	K	K	S	Q	P	H				
Human Six2		A	H	H	K	K	S	Q	P	H				
Nematodes C.elegans		*E*	*E*	*R*	*N*		*S*		*S*	*E*	*H*	*R*		*V*
Nematodes C.elegans	*V*				*V*	*K*	*C*	*K*	*P*	*H*		*K*	*E*	*I*
Nematodes C.briggsae		*E*	*E*	*R*	*N*		*T*		*S*	*E*	*H*	*R*		*V*
Nematodes C.briggsae	*V*				*S*		*C*	*K*	*P*	*H*		*K*	*D*	*V*
Nematodes A.suum		*N*	*D*	*R*	*R*		*N*	*Q*	*P*	*E*				
Nematodes M.chitwoodi	*N*	*P*	*E*	*R*	*N*	*T*	*S*	*N*	*P*	*L*	*H*			
Jellyfish A.Aurita		*N*	*E*	*S*	*N*	*N*	*N*	*N*	*I*	*S*	*S*	*R*		
Jellyfish C.radiatum		*S*	*D*	*Y*	*T*	*H*	*N*	*N*	*Q*		*Y*	*R*		
Jellyfish P.carnea		*N*	*D*	*Y*	*A*	*H*	*N*	*N*	*Q*	*N*	*S*	*R*		
Platyhelminth D.japonica		*P*	*Q*		*T*		*S*	*T*	*P*	*H*				
Platyhelminth D.japonica		*P*	*Q*		*T*		*S*	*T*	*P*	*H*				
Sponge Halichona		*A*	*Q*		*K*		*L*	*S*	*P*	*E*	*S*	*K*		

Table 4.2.2.2: Conservation through evolutionary time of key amino acids of *D. melanogaster* Six domain and homeodomain of So protein. The first row refers to the numbers in figure 4.2.1.1. The second row shows the amino acids that were in bold in figure 4.2.1.1 in So. Their respective numbers correspond to their position with regards to the whole So protein. The following rows shows the amino acids at those positions in all the species previously used in the sequence comparisons. The empty cells indicate that the amino acid is identical to the one in *D. melanogaster* So sequence. The letters in the other cells show which amino acid is present. The colours have been added to indicate what characteristic across time has been conserved in the amino acid substitutions (red for aromatic, pink for aliphatic, green for tiny, blue for acidic, dark blue for basic and no colour for other). The vertical line indicates where the Six domain ends and where the homeodomain begins.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Optix	E^{43}	V^{67}	E^{50}	L^{76}	N^{77}	A^{100}	N^{104}	K^{106}	K^{109}	A^{110}	S^{111}	Q^{177}	T^{184}	P^{197}
D.yakuba														
D.erecta														
D.ananassae														
D.pseudoobscura														
D.mojavensis														
D.virilis														
Anopheles gambiae	A			D	R	S	R			T				
Apis mellifera	A			N	Q		R			D			G	
Mouse Six1	A		A	N	K	H				E			S	
Human Six1	A		A	N	K	H				E			S	
Mouse Six2	A		A	N	K	H				E			S	
Human Six2	A		A	N	K	H				E			S	
Nematodes C.elegans	*V*	*P*		*A*	*G*				*P*	*K*	*Y*	*K*	*P*	*Q*
Nematodes C.briggsae	*V*	*P*		*S*	*A*	*S*			*H*	*E*	*Y*	*K*	*P*	*Q*
Nematodes H.glycines	*L*	*L*	*N*	*A*	*M*									
Nematode P.Pacificus	*S*	*R*		*G*	*N*	*T*	*G*		*A*	*D*	*C*			
Jellyfish C.radiatum	*V*	*S*	*L*	*N*	*T*	*Y*	*H*	*R*		*K*			*S*	
Jellyfish P.carnea	*S*	*N*	*L*	*N*	*S*	*Y*	*H*	*R*		*K*			*S*	
Jellyfish H. Magnipapillata	*I*	*N*	*L*	*N*		*F*	*H*	*R*		*K*	*F*	*E*	*S*	
Jellyfish H. Magnipapillata	*I*	*N*	*L*	*N*		*F*	*H*	*R*		*K*	*F*	*E*	*S*	
Platyhelminth G.tigrina	*T*	*S*	*S*	*S*	*R*	*N*	*K*	*R*				*D*	*S*	

Table 4.2.2.3: Conservation through evolutionary time of key amino acids of *D. melanogaster* Six domain and homeodomain of Optix protein. The first row refers to the numbers in figure 4.2.1.1. The second row shows the amino acids that were in bold in figure 4.2.1.1 in Optix. Their respective numbers correspond to their position with regards to the whole Optix protein. The following rows shows the amino acids at those positions in all the species previously used in the sequence comparisons. The empty cells indicate that the amino acid is identical to the one in *D. melanogaster* Optix sequence. The letters in the other cells show which amino acid is present. The colours have been added to indicate what characteristic across time has been conserved in the amino acid substitutions (red for aromatic, pink for aliphatic, green for tiny, blue for acidic, dark blue for basic and no colour for other). The vertical line indicates where the Six domain ends and where the homeodomain begins.

4.3 Discussion

I compared the three *D. melanogaster* paralogues and identified fourteen amino acids specific to each subfamily. I showed that these amino acids either remain identical across orthologues ranging from poriferans to mammals or that their properties are rather usually conserved when amino acids are substituted. I propose that the degree of conservation of these amino acids suggests a functional importance. Of these amino acids, I suggest that four in particular (three in the Six domain and one in the homeodomain) may be of critical importance (table 4.3.1). The three amino acids in the Six domain that I selected are adjacent to one another. I suggest that this tripeptide may be of functional importance since it is different in all three Six proteins and its properties are different in all three. The last amino acid I selected is the one that remains identical in the most species to the one present in *D. melanogaster* overall (tables 4.2.2.1-4.2.2.3). I suggest that the difference of amino acid at that position in the three subfamilies together with the conservation of these amino acids throughout Metazoa conveys functional importance.

In the following chapters I will discuss the conservation and divergence of function of the Six family of homeodomain transcription factors. I propose that the amino acids I selected in this study have a functional importance. These assumptions remain to be tested.

In addition, in this study I focused solely on the fourteen amino acids whose properties were not conserved and were different in all three proteins. These represent 8% of the amino acids in the Six domain and the homeodomain. 48.5% of the amino acids are identical in all three proteins and 32.5% amino acids are identical two by two (table 4.3.2). These may also be of functional importance and contribute

to the difference in function between two of the three subfamilies. I carried out pairwise comparisons and showed that the percentages of identity between the Six domain and the homeodmain corroborate these suggestions (4.3.3).

Future work may involve identifying features of putative functional conservation between two families at a time. These assumptions also need to be tested. This may involve domain swaps and mutagenesis of key amino acids.

D-Six4	I^{249}	K^{250}	Y^{251}	L^{336}
sine oculis	A^{170}	Q^{171}	N^{172}	T^{258}
Optix	K^{109}	A^{110}	S^{111}	P^{197}

Table 4.3.1: The four amino acids of putative functional importance that I selected. The cells in purple are the amino acids in the Six domain, the cells coloured in green are the amino acids in the homeodomain

	Amino acids unique to **Optix** (identical in sine oculis and D-Six4)	Amino acids unique to **sine oculis** (identical in Optix and D-Six4)	Amino acids unique to **D-Six4** (identical in sine oculis and Optix)	Amino acids different in all three proteins (19 show conserved properties)	Amino acids identical in all three proteins	Total
Six domain and homeodomain	*24*	*10*	*22*	*33*	*84*	*173*
Six domain and homeodomain	14%	5.8%	12.7%	19%	48.5%	100%

Table 4.3.2: The number of amino acids identical in the Six domain and the homeodomain in two or three of the Six proteins and the corresponding percentage are shown alongside the number and corresponding percentage of amino acids different in all three proteins.

Percentage of identity	Six domain	Homeo domain
D-Six4 versus sine oculis	59%	72%
D-Six4 versus Optix	53%	59%
sine oculis versus Optix	58%	68%

Table 4.3.3: Percentages of identity between the Six domain and the homeodomains of the three *Drosophila melanogaster* Six proteins.

Chapter 5

Functional conservation and divergence between *Drosophila* paralogues

Functional conservation and divergence between Drosophila paralogues

5.1 Introduction

So and Optix have different expression patterns to D-Six4 (Chapter 1). So and Optix are most well-known for their expression in the head whereas D-Six4, although also expressed in the head, principally drives the patterning of the mesoderm (Chapter 1). So is required for the development of the entire visual system and Optix too has important functions in eye development (Chapter 1). D-Six4 is involved in muscle and gonad development (Chapter 1). These differences are corroborated by differences at the level of the protein sequences. Indeed, the protein sequences of each of these three proteins display amino acid specificities relative to each subfamily (Chapter 4).

D-Six4^{289} is a likely null mutant (Chapter 1). Endogenous So and Optix do not compensate for, at least part of, D-Six4 function resulting in severe muscle and gonad phenotype (Chapter 1). In this chapter, I explore the ability of So and Optix to rescue aspects of the *D-Six4*^{289} mutant phenotype. I hypothesised that even if expressed at the same time and same place as D-Six4, in a *D-Six4*^{289} mutant, neither So nor Optix would be able to rescue parts of the mutant phenotype. I hypothesised therefore that beyond the differences of expression patterns, the three Six proteins have genuinely functionally diverged.

Eya is expressed both in the head and the mesoderm (Chapter 1) and is a cofactor of So in the development of the eye. Eya is also coexpressed with D-Six4 in the mesoderm (Chapter 1). This suggests that Eya and D-Six4 might interact. I hypothesised that even if So were to interact with Eya and assuming that the So/Eya

and D-Six4/Eya interactions were similar, it would be insufficient to rescue aspects of the *D-Six4*^{289} mutant phenotype because of other factors involved including *tinman* and *Mef 2* in the muscle and *Hmgcr* in the gonads (Chapter 1).

Ectopic expression of Optix alone is sufficient to induce eyes in the wings and halteres (Chapter 1). It is unclear whether this is the result of the ability of Optix to act on its own or whether it is the result of its potential to interact with different cofactors to carry out the same function. Until recently, this was thought to be different to So which was thought to only be able to induce ectopic eyes through an Eya-dependent mechanism (Chapter 1). I hypothesised that assuming that the ability of Optix to induce ectopic eyes was the result of its ability to act on its own then this might suggest a capability to rescue aspects of the *D-Six4*^{289} mutant phenotype as its functioning may not be dependent on other factors. However, according to the classification of the Six proteins into three subfamilies, proteins within a subfamily are more closely related to one another than they are to the other proteins within the same species (Seo et al., 1999). Additionally, Six3 in mouse acts as a transcriptional repressor whereas members of the Six1/Six2 and Six4/Six5 subfamilies act as transcriptional activators (Chapter 1). It is therefore possible that Optix acts also as a transcriptional repressor and that So and D-Six4 act as transcriptional activators, making it unlikely for Optix to rescue aspects of the *D-Six4*^{289} mutant phenotype.

I assessed the respective ability of D-Six4, So and Optix to rescue aspects of the *D-Six4*^{289} mutant phenotype in the mesoderm by carrying out *in vivo* developmental genetic assays. One of my aims was to establish to what extent the Six family of proteins in *D. melanogaster* were functionally conserved, if at all.

Eya is both involved in the development of muscle and gonads and mesodermal phenotypes of *eya* mutants closely resemble those of *D-Six4* for SGP and muscle development (Chapter 1). In *D-Six4*^{289} mutants, endogenous expression levels of Eya are insufficient to rescue the mutant phenotype. I overexpressed Eya in order to assess whether Eya, if present in high enough levels, is able to compensate for loss of D-Six4 function in the mesoderm. My aim was to further understand the role of Eya together with D-Six4 and their patterning of the mesoderm.

5.2 Results

The results described in this chapter are based on experiments in which I analysed the effect of gene misexpression in the mesoderm of *D-Six4* mutant embryos. I collected, histochemically stained and analysed embryos through confocal microscopy. The effect of expression of the D-Six4, So, Optix and Eya proteins in the mesoderm of heterozygous embryos was also analysed. I will discuss these results in turn.

5.2.1 How did I express the Six proteins and Eya in the mesoderm?

Prior to the start of my PhD, five transformant fly lines were available containing five different p-elements. Four of these were an *E.coli* plasmid with a yeast upstream activating sequence fused to *so*, *eya*, *Optix* and *D-Six4* respectively (Clark et al., 2006; Kenyon et al., 2005; Pignoni et al., 1997). The fifth transformant line had a p-element which was an *E.coli* plasmid containing the mesodermal promoter of *twist* fused to GAL 4 (Greig and Akam, 1993).

When I started my PhD, the transformant lines had different genetic backgrounds and the p-elements had inserted on different chromosomes (Ivan Clark, personal communication) (Clark et al., 2006; Kenyon et al., 2005; Pignoni et al., 1997), (Greig and Akam, 1993). I consequently had to carry out a number of crosses before I was able to use these lines. I discuss these in turn.

5.2.1.1 What crosses did I have to carry out with the tranformant lines containing the p-elements [UAS-*D-Six4*], [UAS-*so*] and [*twist*-GAL4] respectively?

As I started my PhD, both the transformant lines containing p[*twist*-GAL4] and the p[UAS-*D-Six4*] were in a balanced *D-Six4*^{289} mutant background. *D-Six4*^{289} homozygotes are embryonic lethal. The *D-Six4*^{289} stock is kept over a balancer, a genotypically wild type chromosome that has nested inversions so as to prevent recombination with the other chromosome which would result in the loss of the mutation. p[*twist*-GAL4] was homozygous on the 2nd chromosome and on the third chromosome, the *D-Six4*^{289} mutation was balanced with a third multiple balancer (TM3) containing a p[Ubx-lacZ] inserted p-element used for the identification of putative rescued embryos (discussed later) (Ivan Clark, personal communication). p[UAS-*D-Six4*] was on the third chromosome carrying the *D-Six4*^{289} mutation and balanced with the same balancer, TM3 (Ivan Clark, personal communication). Consequently, no further cross was required prior to using these transformant lines.

Similarly, the transformant line containing the [UAS-*so*] p-element was also in a balanced *D-Six4*^{289} mutant background, on the third chromosome carrying the *D-Six4*^{289} mutation (Ivan Clark, personal communication). However it was balanced with a different third multiple balancer (TM6) (Ivan Clark, personal communication).

Therefore I carried out a balancer swap before using this transformant line (figure 5.2.1.1.1).

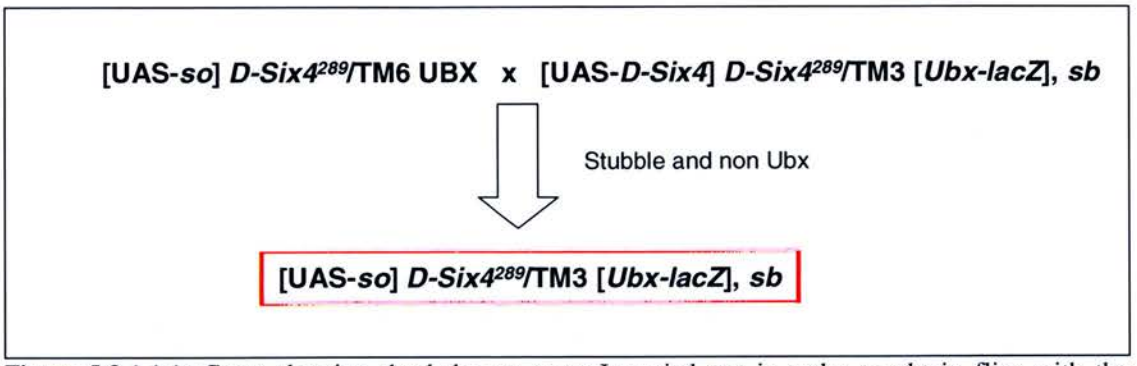


Figure 5.2.1.1.1: Cross showing the balancer swap I carried out in order to obtain flies with the genotype in the red box. Those flies displayed stubble bristles and wild-types halteres.

5.2.1.2 What crosses did I have to carry out with the tranformant lines containing the p-elements [UAS-*Optix*] and [UAS-*eya*] respectively?

As I started my PhD, p[UAS-*Optix*] was homozygous on the third chromosome but was in a wild-type background (Ivan Clark, personal communication). I therefore had to carry a number of crosses to change the background to a *D-Six4*^{289} mutant one and balanced with TM3 (figure 5.2.1.2.1). One of these crosses involved a recombination event allowing for p[UAS-*Optix*] to be on a chromosome carrying the *D-Six4*^{289} mutation (figure 5.2.1.2.1).

As I started my PhD, p[UAS-*eya*] was homozygous on the second chromosome, in a *D-Six4*^{289} mutant background but balanced with the third multiplier TM6 (Ivan Clark, personal communication). I therefore carried out a balancer swap. This however required more crosses than was the case with the p[UAS-*so*] balancer swap, as indeed, p[UAS-*eya*] was on the second chromosome (figure 5.2.1.2.2).

Once these crosses had been carried out, I was able to use both these transformant lines in order to express the proteins in the mesoderm.

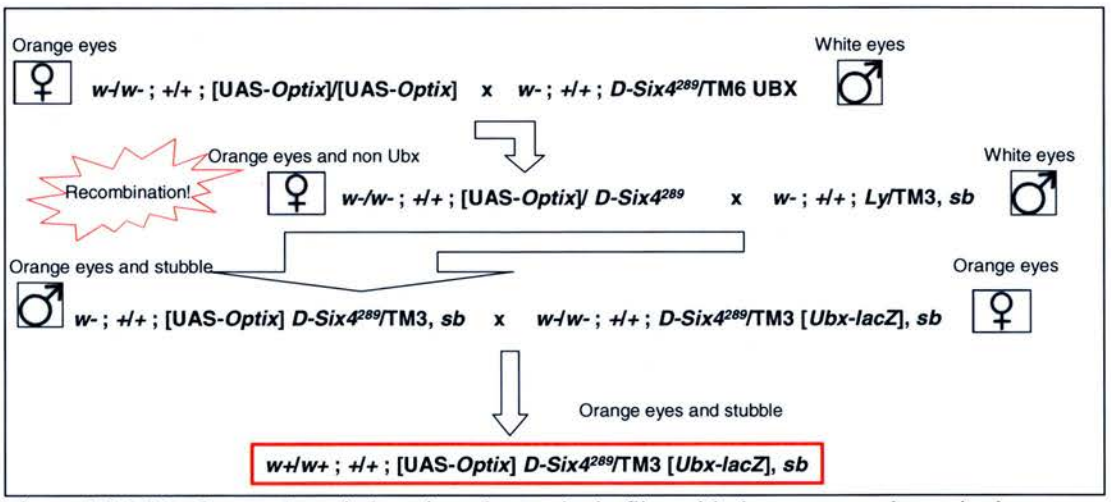


Figure 5.2.1.2.1: Crosses I carried out in order to obtain flies with the genotype shown in the red box. These were crossed with males of the same genotype and kept as a stock.

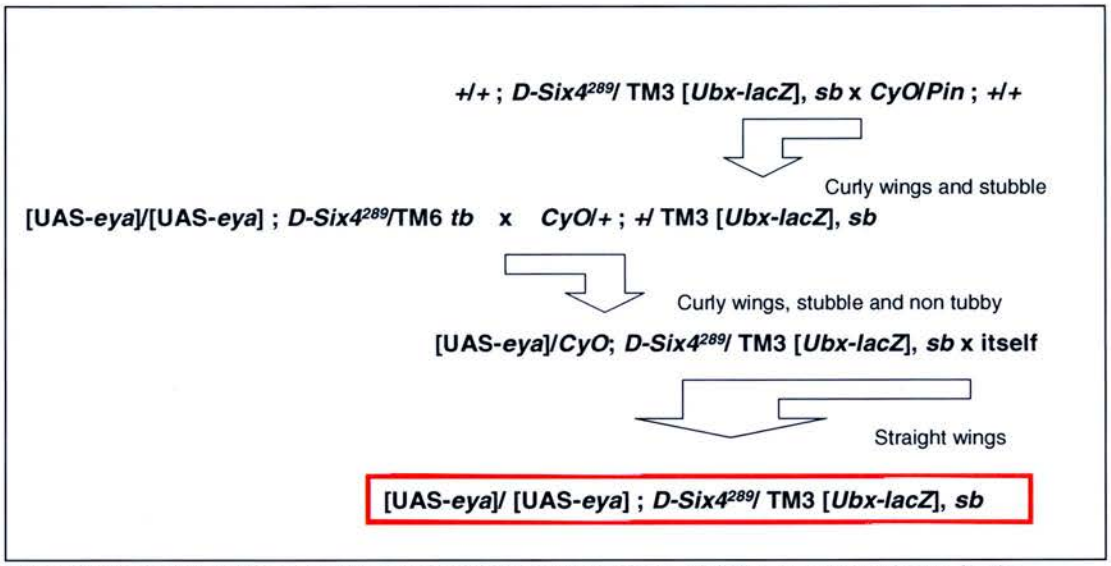


Figure 5.2.1.2.2: Crosses I carried out in order to obtain flies with the genotype shown in the red box. These were crossed with one another and kept as a stock.

5.2.1.3 What cross did I carry out in order to express the proteins in the mesoderm?

Following the experimental approach of Brand and Perrimon, I crossed the transformant lines containing the [UAS-*D-Six4*], [UAS-*so*], [UAS-*Optix*] and [UAS-*eya*] p-elements, respectively, with the transformant line containing the [*twist*-GAL4] p-element (figure 5.2.1.3.1) (Brand and Perrimon, 1993). In the resulting progeny, the yeast upstream activating sequence when bound by GAL4, driven in the mesoderm by an enhancer from *twist,* enables the transcription and translation of the fused gene in the mesoderm (figure 5.2.1.3.1).

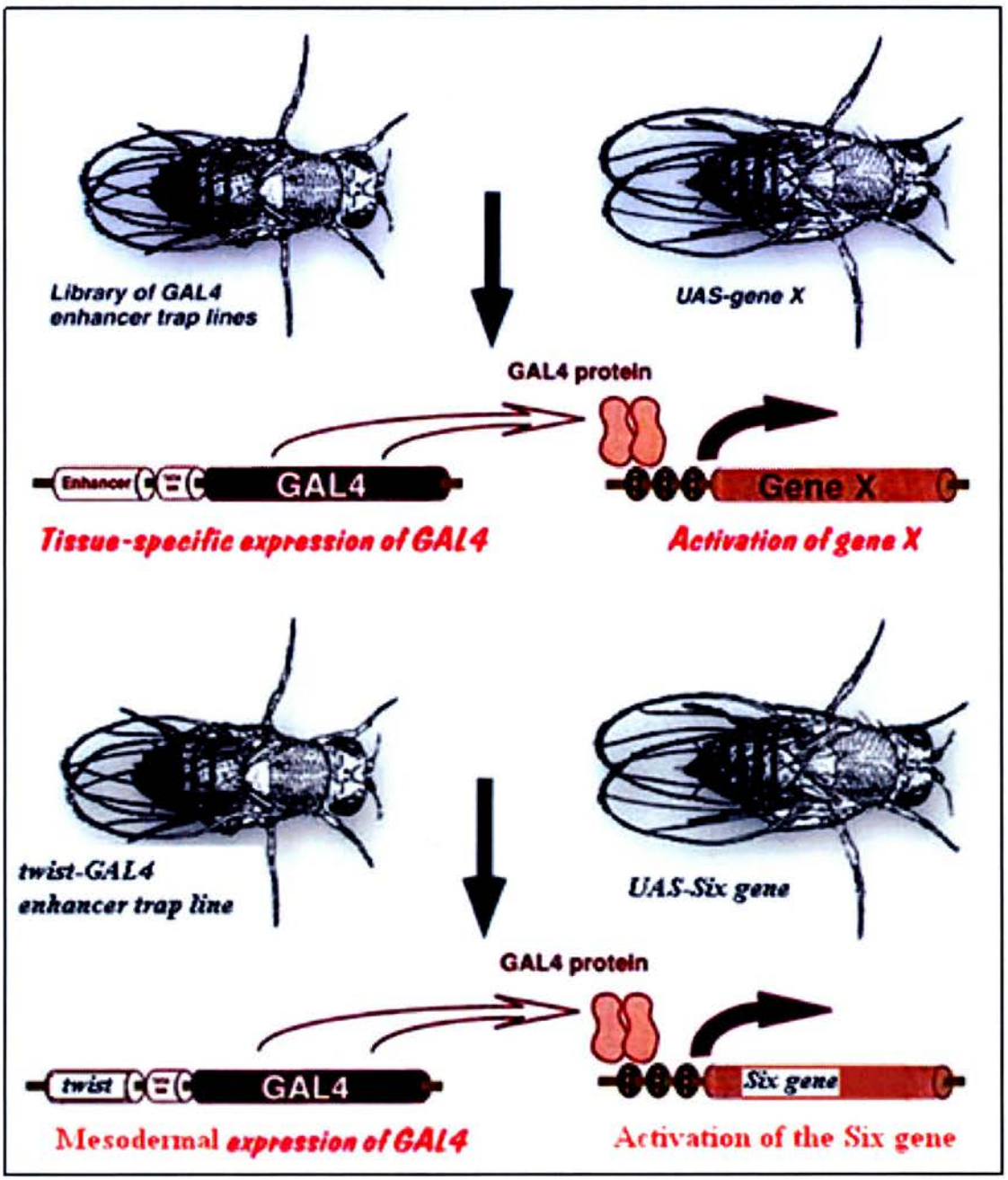


Figure 5.2.1.3.1: Directed gene expression in *Drosophila*. To generate transgenic lines expressing GAL4 in numerous cell- and tissue-specific patterns, the GAL4 gene is inserted randomly into the genome, driving GAL4 expression from numerous different genomic enhancers (Brand and Perrimon, 1993). A GAL4-dependent target gene can then be constructed by subcloning any sequence behind GAL4 binding sites, in these experiments *Six* genes and *eya* (Brand and Perrimon, 1993). The target gene is silent in the absence of GAL4 (Brand and Perrimon, 1993). To activate the target gene in a cell- or tissue-specific pattern, flies carrying the target (*UAS-Gene X*) are crossed to flies expressing GAL4 (Enhancer Trap GAL4) (Brand and Perrimon, 1993). In the progeny of this cross, it is possible to activate *UAS-Gene X* in cells where GAL4 is expressed and to observe the effect of this directed misexpression on development (Brand and Perrimon, 1993).

5.2.2 How did I collect and histochemically stain the embryos?

I carried out the cross previously described (5.2.1.3.1) and collected and stained resulting embryos (chapter 8). *D-Six4*^{289} homozygous mutant embryos display a disrupted muscle structure, an absence of SGPs and scattered germ cells. In order to assess putative rescue of these features by the expression of the different proteins, I used primary antibodies anti-Myosin, anti-eya and anti-vasa to stain for the muscle structure, the SGPs and the germ cells respectively.

Four genotypes result from the cross described previously (figure 5.2.2.1). In order to identify embryonically the *D-Six4*^{289} homozygotes, the balancer I used had an inserted p-element, p[Ubx-*lacZ*]. Embryos whose genotype had at least one copy of this balancer, three out of four genotypes, stained for beta-galactosidase. *D-Six4*^{289} homozygotes did not. I therefore used the primary antibody anti-beta-galactosidase to identify *D-Six4*^{289} homozygotes.

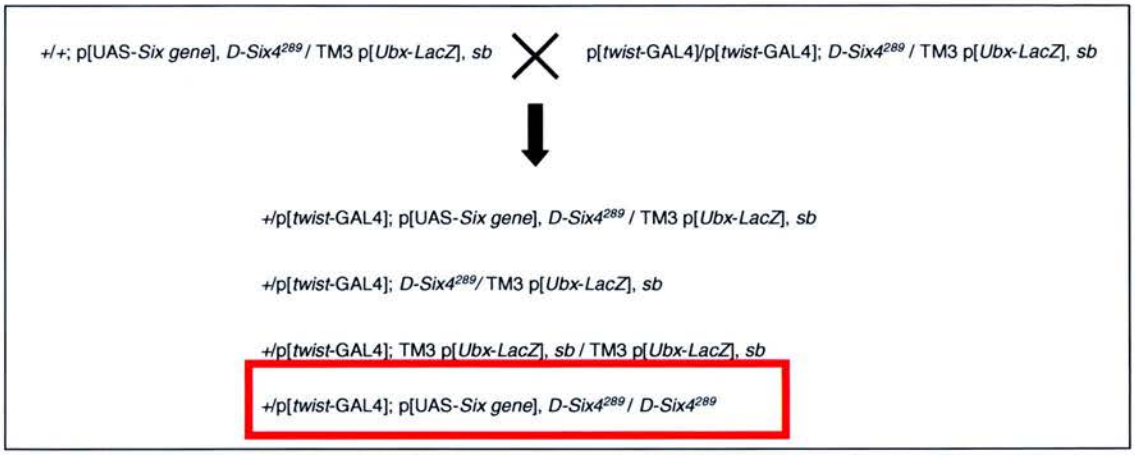


Figure 5.2.2.1 Four genotypes result from the cross carried out in these experiments. Three genotypes contain the balancer chromosome and stain for beta-galactosidase and one genotype does not stain for beta-galactosidase (red box) allowing the identification of *D-Six4*^{289} homozygotes. P[UAS-*Six* gene] is short for P[UAS-*so*], P[UAS-*D-Six4*], P[UAS-*Optix*] and P[UAS-*eya*].

5.2.3 How did I analyse the embryos?

I observed embryos using a confocal microscope. In wild-type embryos, D-Six4 mRNA is restricted to ventral and lateral mesoderm by stage 10 driving the development of the muscle (Chapter 1). However, the *D-Six4*^{289} mutant phenotype is visible using the primary antibody anti-Myosin from stage 13 onwards, after germ band retraction, when the developed (or in this case undeveloped) muscles express Myosin. The muscle arrangements at stages 13-15 are consequently an indicator of D-Six4 function. I therefore looked at stage 13-15 embryos that did not stain for anti-Beta-galactosidase and assessed the ventral and lateral muscle phenotypes.

When I started my PhD, I used published figures of stages extensively in order to stage the embryos I was looking at. I carried out visual comparisons of the external shape of the embryo (Weigmann et al., 2003). As time went on, during meetings and discussions, I would explain my work referring to embryos at specific stages. Not being contradicted by my fellow colleagues and supervisors, I concluded that I was staging my embryos correctly and did not seek a different way of doing this. This was definitely an oversight. I should have used precise criteria and maybe carried out co-stainings with markers of specific stages in order to assess specifically the stages I was looking at. Indeed, the muscle phenotype changes considerably between stages 13-15. Had I done this, it may have affected my interpretation of the results as I discuss later.

In wild-type embryos, at stage 13, D-Six4 is expressed weakly in the SGPs in parasegments 10-12 and then by stage 15 is expressed strongly in the SGPs after they have coalesced with the migrating germ cells (Chapter 1). Failure of D-Six4 function in *D-Six4*^{289} results in SGPs cell death (Chapter 1). I looked at stage 15-17 embryos

that again did not stain for anti-beta-galactosidase as at those stages in wild-type embryos SGPs express Eya and should therefore be visible using the primary antibody anti-eya. Again, as discussed previously, staging was not carried out thoroughly enough. Also, staining with the anti-eya antibody was not straight forward. Two problems arose. The anti-eya antibody does not stain as efficiently as for example the anti-vasa antibody, discussed later, and most of the time, staining with anti-eya was a means of assessing an absence of SGPs. This meant that it was difficult to assess whether absence of visible SGPs was the result of their genuine absence or of poor staining. This question should have been easily answered by comparing with the controls but even in a wild-type embryo, stained only with anti-eya, it is sometimes difficult to assess whether SGPs are present. To address this, I co-stained embryos together with anti-vasa that is expressed in germ cells in order to at least in the controls try to increase the chances of determining whether SGPs are present. This resulted in a second problem: using the confocal microscope, the anti-vasa antibody stained so efficiently that the signal for the anti-eya antibody was swamped by the anti-vasa signal. Despite many attempts to address this issue, it was not resolved. Consequently, I assessed putative rescue of aspects of the *D-Six4*^{289} mutant phenotype and rescue of D-Six4 function by looking at germ cells only.

Although D-Six4 is not expressed in the germ cells in wild-type embryos, *D-Six4*^{289} mutants result in the scattering of germ cells possibly because of the absence of SGPs failing to attract the germ cells into the gonad (Chapter1). Consequently, the scattering of the germ cells, although only an indirect effect of the *D-Six4*^{289} mutation, is an indicator of D-Six4 function. Hence I assessed the germ cells phenotype in stage 15-17 embryos that did not stain for beta-galactosidase. I applied

specific criteria to assess whether the phenotypes observed indicated putative rescue of the mutant phenotype. I discuss these below.

5.2.3.1 What are the criteria I used to assess putative rescue of the muscle phenotype?

The *D-Six4*^{289} mutant phenotype was previously characterised and I used these criteria as a means of classifying the muscle phenotypes I observed when expressing D-Six4, So, Optix and Eya in the mesoderm.

I classed embryos as fully rescued when they displayed the regular pattern of myotubes in the lateral and ventral somatic muscles observed in the wild type embryos and when they were thus indistinguishable from wild type embryos (figure 5.2.3.1.1A). I classed no embryo as partial rescue. Indeed, in the *D-Six4*^{289} homozygous mutant, embryos have strongly disrupted lateral and somatic muscles and some muscles appear to be entirely missing. However because the number and location of these muscles varies between segments and between embryos, it is very difficult to assess whether there is partial rescue of these muscles when expressing the exogenous proteins. Not classifying any embryos as partial rescue may have been a mistake. Indeed, it seems unlikely that rescue is either "all or none". A more careful analysis would have perhaps enabled me to discern intermediate phenotypes. As discussed earlier, my staging of embryos may have been imprecise. It may be that had I staged my embryos more accurately, I would have been able to discern partial phenotypes. This would also have been possible by staining with muscle specific anti-bodies, as discussed later.

I classed embryos as not rescued if they displayed a phenotype indistinguishable from the previously characterised mutant phenotype (figure

5.2.3.1.1B). As just mentioned, the *D-Six4*^{289} mutant muscle phenotype displays a variation in the number and location of the muscles affected and therefore there isn't one muscle mutant phenotype as such but several. This adds an additional difficulty in trying to assess rescue of the mutant phenotype. Since previous description of the *D-Six4*^{289} mutant muscle phenotype was summarised as a disruption of the ventral and lateral muscles (Clark et al., 2006; Kirby et al., 2001), embryos displaying disrupted ventral and lateral muscles were classed as non rescued. More specific characterisation of the mutant phenotype will certainly help resolve these difficulties. Finally, I classed embryos as worsened when they displayed a more severe phenotype than the characteristic *D-Six4*^{289} mutant one. This phenotype was recognisable as the observed disruption of the muscles was significantly greater than that observed in the *D-Six4*^{289} mutant embryos (discussed later). Rescue of D-Six4 function was then assessed as a proportion of the number of mutant embryos in which full rescue was observed.

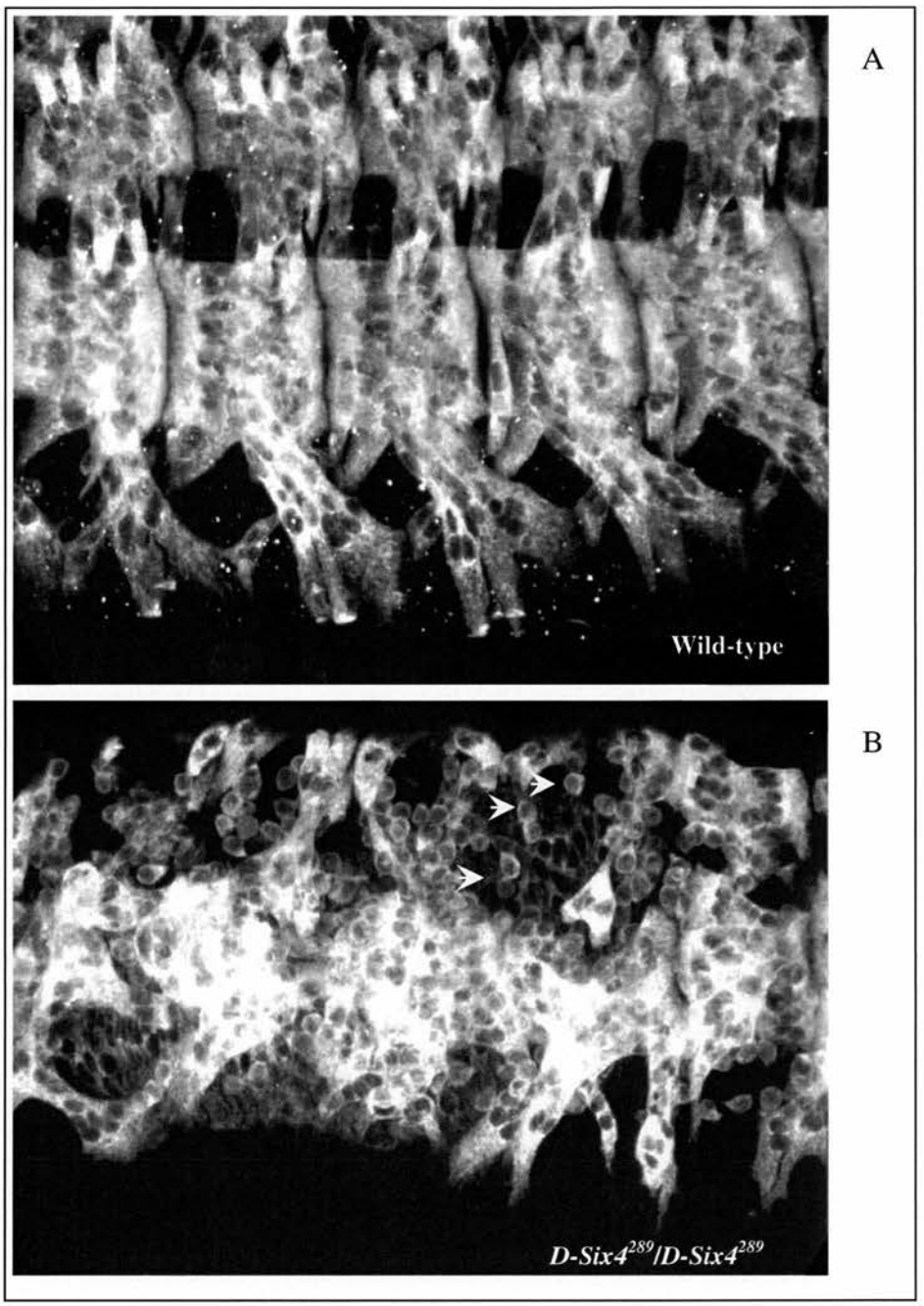


Figure 5.2.3.1.1. A: Mid ventro-lateral section of a wild-type embryo, stained for anti-Myosin showing the regular arrangement of syncytial myotubes. B: Mid ventro-lateral section of a homozygous *D-Six4*^{289} embryo stained for anti-Myosin, where muscles are highly disorganized; some are missing, and there are many unfused myocytes (arrowheads).

5.2.3.2 What are the criteria I used to assess putative rescue of the gonad phenotype?

As for the muscle, the *D-Six4*^{289} mutant phenotype in the gonad was previously characterised and again I used these criteria as a means of classifying the gonad phenotypes I observed while expressing D-Six4, So, Optix and Eya in the mesoderm. Additionally, I refined these criteria more specifically.

I classed embryos as fully coalesced when the germ cells were coalesced into two tightly formed gonads indistinguishable from a wild-type gonad (figure 5.2.3.2.1). I classed embryos as scattered when germ cells were scattered. The meaning of scatter here is critical. Previous characterisation of the mutant phenotype either defined germ cells as coalesced or not. Here I define scattering of germ cells as dispersed across the posterior of the embryo displaying no clustering of more than three germ cells (figure 5.2.3.2.2A). I classed embryos as partially scattered when germ cells displayed one or more clusters of more than three germ cells. The remainder of the germ cells were dispersed across the posterior of the embryo (figure 5.2.3.2.2B). Variation in phenotypes classed as scattered and partially scattered was not assessed beyond the criteria mentioned above. However a more precise classification of these phenotypes could have included the number of germ cells scattered and the distances between them.

As discussed previously, fully coalesced germ cells are not observed in the *D-Six4*^{289} mutant. However, embryos displaying both scatter and partial scatter of germ cells are observed in some of the embryos expressing exogenous proteins as well as in some of the mutant embryos not expressing exogenous proteins (discussed later). In the *D-Six4*^{289} mutant, there is no reported expression of D-Six4. This partial

scatter phenotype can be explained in several ways. It may be part of the mutant phenotype which therefore would be more heterogeneous than previously described (Kirby et al., 2001). An alternative is that the *D-Six4*^{289} mutation is not a null mutant as previously thought and some basal level of D-Six4 is being expressed allowing for the clustering of a few germ cells with one or two SGPs. The protein level expressed is too low for detection but sufficient to give rise to this phenotype. Instead of a null mutant, the *D-Six4*^{289} mutation would be a more severe hypomorph than the *D-Six4*^{131} mutant. It may be that some SGPs form in complete absence of D-Six4 function. As discussed later, recent work has highlighted different functional pathways involved in the formation of the gonads. It may be that D-Six4 function is not required in one of these pathways. Another alternative is that there is some sort of redundancy on the behalf of other factors present in the mesoderm that are able in certain circumstances to carry out part of D-Six4 function. It may be argued that these phenotypes are the result of the presence of the p-elements. However, since these phenotypes were observed in every one of the controls carried out all carrying different p-elements having inserted in different places of the genome (discussed later), I propose that these phenotypes are a genuine phenotype of the *D-Six4*^{289} mutation. Analysis of the results obtained in this PhD may help decipher which of these possibilities is more likely.

Rescue of D-Six4 function was then assessed as a proportion of the number of mutant embryos in which full coalescence was observed. Partial rescue of D-Six4 was assessed only when the percentage of embryos observed displaying partial scatter when expressing the exogenous proteins in the mutant embryos was greater

than the percentage of embryos displaying partial scatter observed in the *D-Six4*^{289} homozygous mutant not expressing any exogenous proteins.

Additionally, I counted the number of germ cells. In the embryos expressing D-Six4, So, Optix and Eya, I only counted the germ cells in a proportion of fully coalesced embryos. In the mutant embryos not expressing any exogenous protein, I counted the scattered germ cells and divided that number by half to get an estimate of the number of germ cells per uncoalesced gonad.

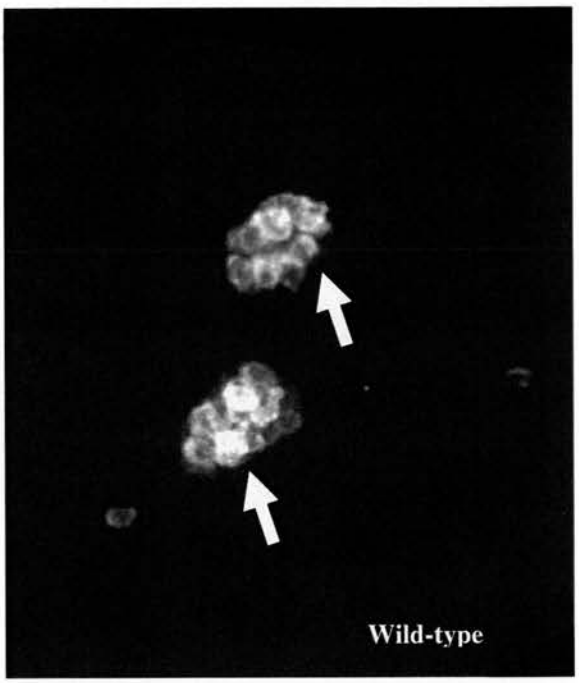


Figure 5.2.3.2.1: Posterior section of a wild-type embryo stained for anti-vasa showing stained germ cells coalesced to form two gonads (arrows).

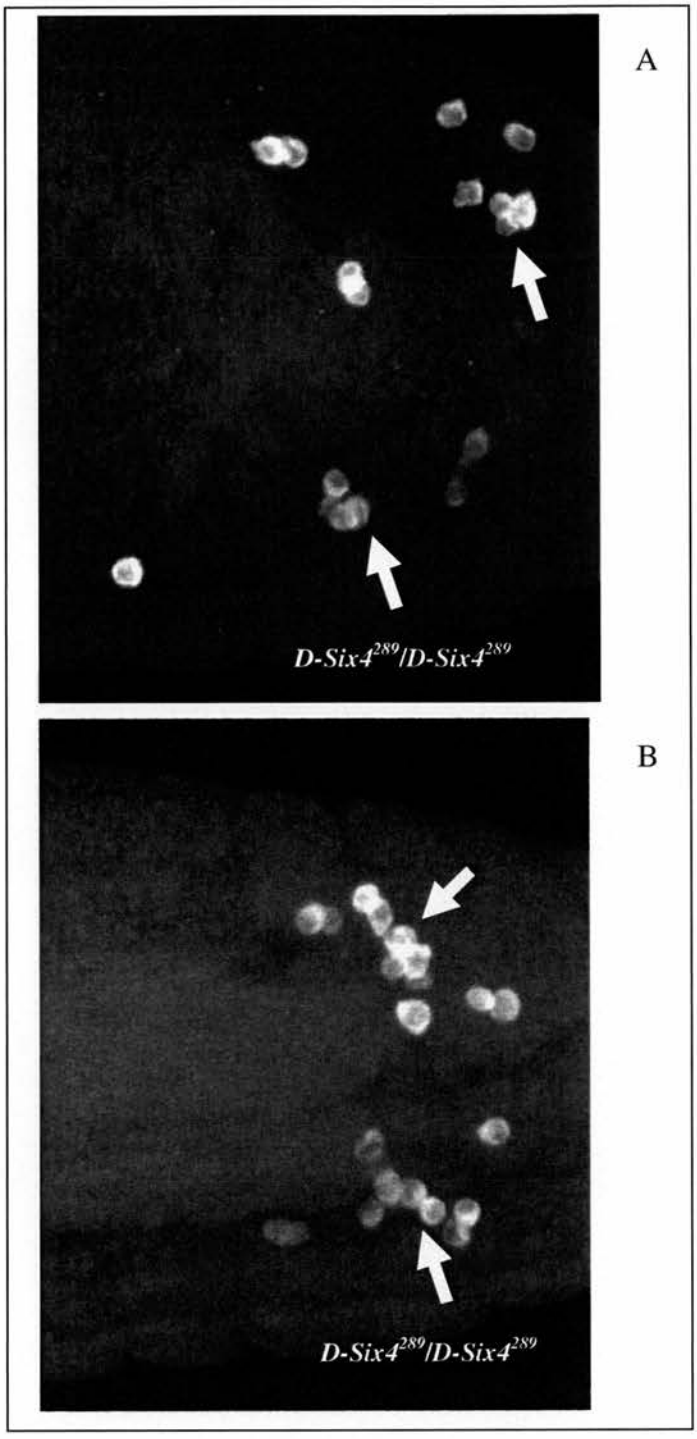


Figure 5.2.3.2.2 A and B: Posterior sections of *D-Six4*^{289} homozygous mutants stained for anti-vasa. A: germ cells display clusters of no more than 3 germ cells (arrows) and are therefore scattered across the posterior of the embryo B: germ cells display two clusters of more than 3 germ cells (arrows) and are therefore partially scattered.

5.2.3.3 What controls did I carry out?

I collected, histochemically stained and analysed through confocal microscopy embryos of each parental strain crossed with itself to ensure that the p[UAS-*so*], p[UAS-*D-Six4*], p[UAS-*Optix*] and p[UAS-*eya*] constructs and the p[*twist*-GAL4] construct did not have an effect on the *D-Six4*^{289} mutant phenotype.

5.2.3.4 Was there rescue of the ventral and lateral muscles when expressing the *Six* genes?

As expected, the control experiments I carried out all displayed the previously characterised *D-Six4*^{289} homozygous mutant phenotype: the phenotypes I observed were indistinguishable from the previously characterised mutant phenotypes. I classed these embryos as not rescued. No visible effect of the presence of the p[UAS-*so*], p[UAS-*D-Six4*], p[UAS-*Optix*] and p[UAS-*eya*] constructs and the p[*twist*-GAL4] construct was observed.

Expression of D-Six4, So and Optix in the mesoderm of a *D-Six4*^{289} homozygous mutant embryo results in different muscle phenotypes in different proportions.

5.2.3.4.1 What was the effect of expressing exogenous D-Six4?

Expression of D-Six4 using a *twist*-GAL4 driver results in substantial rescue of the *D-Six4*^{289} homozygous mutant muscle phenotype. I observed embryos displaying a regular pattern of myotubes in the lateral and ventral somatic muscles (figure 5.2.3.4.1.1). In these embryos, no muscles appeared to be missing or disrupted (figure 5.2.3.4.1.1). They were indistinguishable from wild-type embryos (figure 5.2.3.4.1.1). 75% of the embryos assessed (n=53) were thereby classed as fully rescued (figure 5.2.3.4.1.2). The remainder of the embryos, 25%, displayed a mutant phenotype indistinguishable from the previously characterised mutant phenotype. I classed these embryos as not rescued(figure 5.2.3.4.1.3).

In summary, 75% of the *D-Six4*^{289} homozygous mutant embryos show rescue of D-Six4 function when D-Six4 is expressed using a *twist*-GAL4 driver (figure 5.2.3.4.1.2).

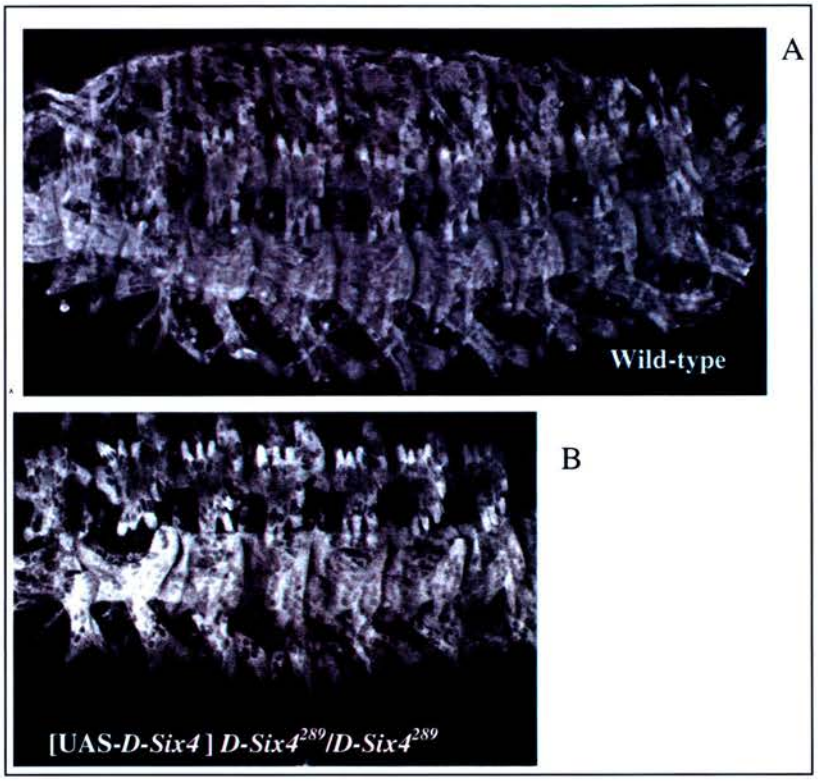


Figure 5.2.3.4.1.1: A: Wild-type embryo stained for anti-Myosin showing a regular array of ventral and lateral muscles. B: Homozygous *D-Six4*^{289} embryo stained for anti-Myosin expressing [UAS-*D-Six4*] in the mesoderm under the control of a *twist*-GAL4 driver displaying a phenotype indistinguishable from wild type (A).

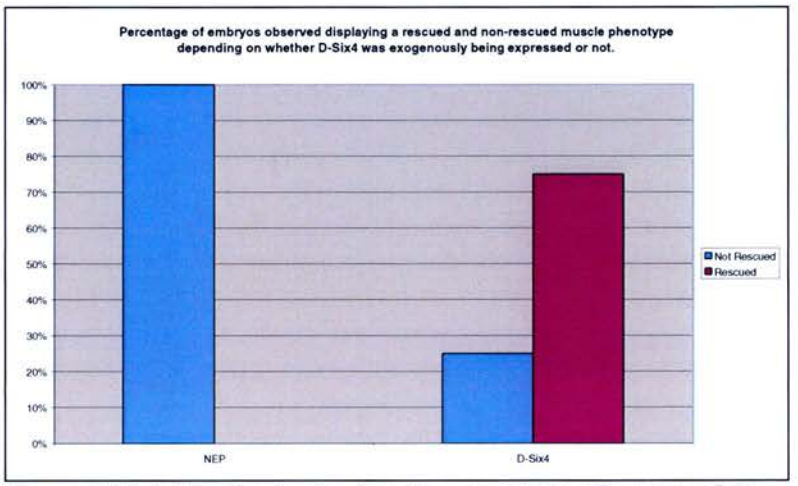


Figure 5.2.3.4.1.2: Graph showing the percentage of rescue of D-Six4 function in the muscle when no exogenous protein is being expressed, (NEP), compared with the percentage of rescue of D-Six4 function in the muscle when expressing [UAS-*D-Six4*] in the mesoderm of a homozygous *D-Six4*^{289} embryo under the control of a *twist*-GAL4 driver (D-Six4).

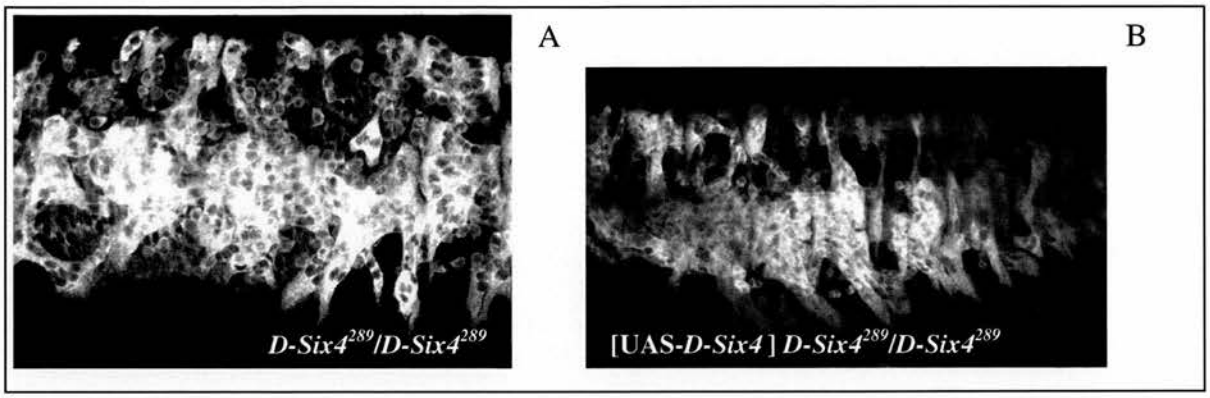


Figure 5.2.3.4.1.3: A: Homozygous *D-Six4*^{289} embryo stained for anti-Myosin displaying a disrupted musculature. B: Homozygous *D-Six4*^{289} embryo expressing [UAS-*D-Six4*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-Myosin displaying a phenotype indistinguishable from the homozygous *D-Six4*^{289} mutant phenotype (A).

5.2.3.4.2 What was the effect of expressing exogenous So and Optix?

Expression of both So and Optix, using a *twist*-GAL4 driver results in rescue of the *D-Six4*^{289} homozygous mutant muscle phenotype. Similarly to the muscle phenotype observed when expressing D-Six4 using a *twist*-GAL4 driver, I observed embryos displaying a regular pattern of myotubes in the lateral and ventral somatic muscles that were indistinguishable from wild-type embryos (figure 5.2.3.4.2.1). I classed these embryos as fully rescued. The percentage of embryos I classed as fully rescued was lower than that observed when expressing D-Six4 using a *twist*-GAL4 driver both in the embryos expressing So, 30% (n=56) and Optix, 8% (n=27) respectively as opposed to D-Six4, 75% (n=53) (figure 5.2.3.4.2.2).

In the remaining 70% of embryos expressing So, the muscle phenotype observed was indistinguishable from the previously characterised mutant phenotype (figure 5.2.3.4.2.3). I consequently classed these embryos as not rescued.

In all of the remaining, 92% of embryos expressing Optix, more ventral muscles were absent (figure 5.2.3.4.2.3). No ventral oblique or ventral longitudinal muscles were present (figure 5.2.3.4.2.3). There appeared to be more unfused myoblasts and fewer elongated unfused founder cells attempting to form myoblasts (figure 5.2.3.4.2.3). The phenotype observed appeared more severe than the previously characterised mutant one. I classed these embryos as worsened.

In summary, 30% and 8% of the *D-Six4*^{289} mutant embryos show rescue of D-Six4 function when So and Optix are expressed respectively using a *twist*-GAL4 driver (figure 5.2.3.4.2.2).

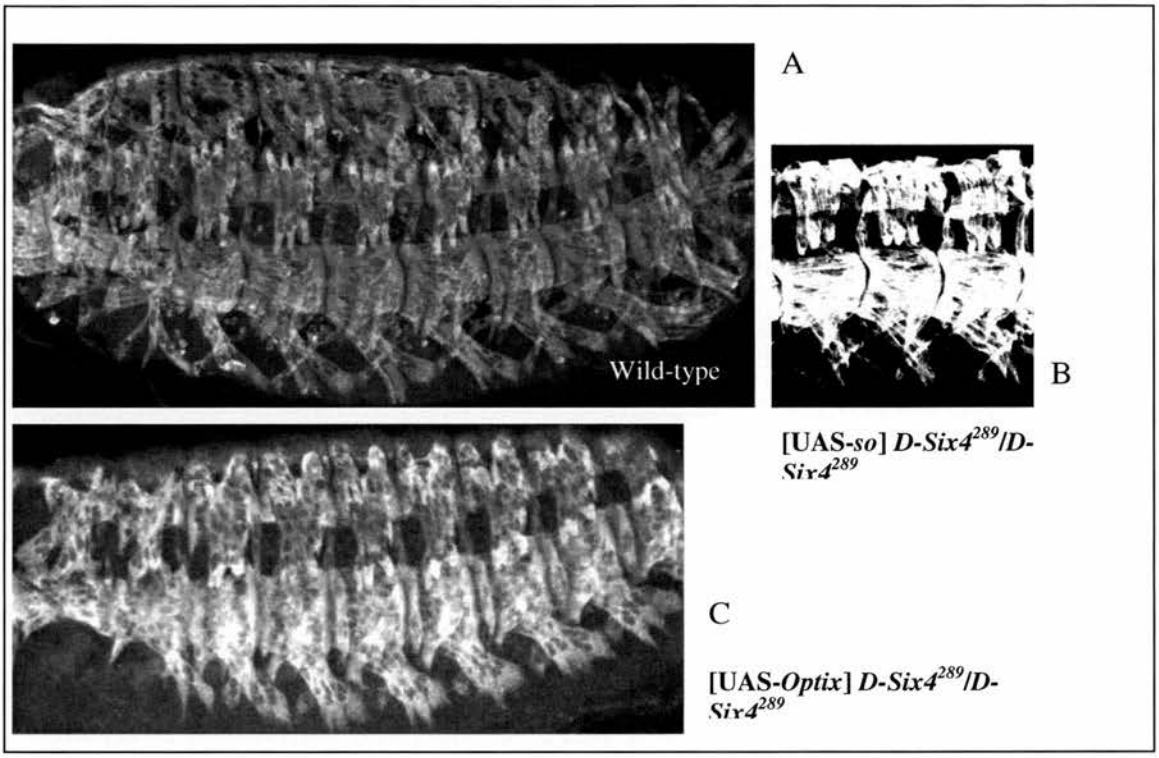


Figure 5.2.3.4.2.1: A: A wild-type embryo stained for anti-Myosin displaying a regular array of muscles. B: A mid ventro-lateral section of a homozygous *D-Six4*^{289} embryo stained for anti-Myosin expressing p[UAS-*so*] in the mesoderm under the control of the *twist*-GAL4 driver. The array of ventral and lateral muscles is indistinguishable from wild-type (A). C: A homozygous *D-Six4*^{289} embryo stained for anti-Myosin expressing p[UAS-*Optix*] in the mesoderm under the control of the *twist*-GAL4 driver. The array of ventral and lateral muscles is indistinguishable from wild-type (A).

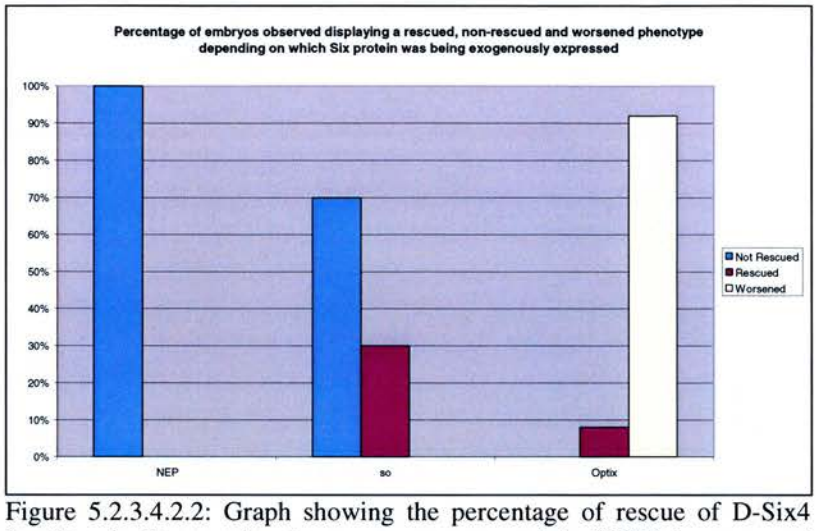


Figure 5.2.3.4.2.2: Graph showing the percentage of rescue of D-Six4 function in the muscle when no exogenous protein (NEP) is expressed compared to the percentage of rescue of D-Six4 function in the muscle when expressing [UAS-*so*], (so), and [UAS-*Optix*], (Optix), in the mesoderm of a homozygous *D-Six4*^{289} embryo under the control of a *twist*-GAL4 driver.

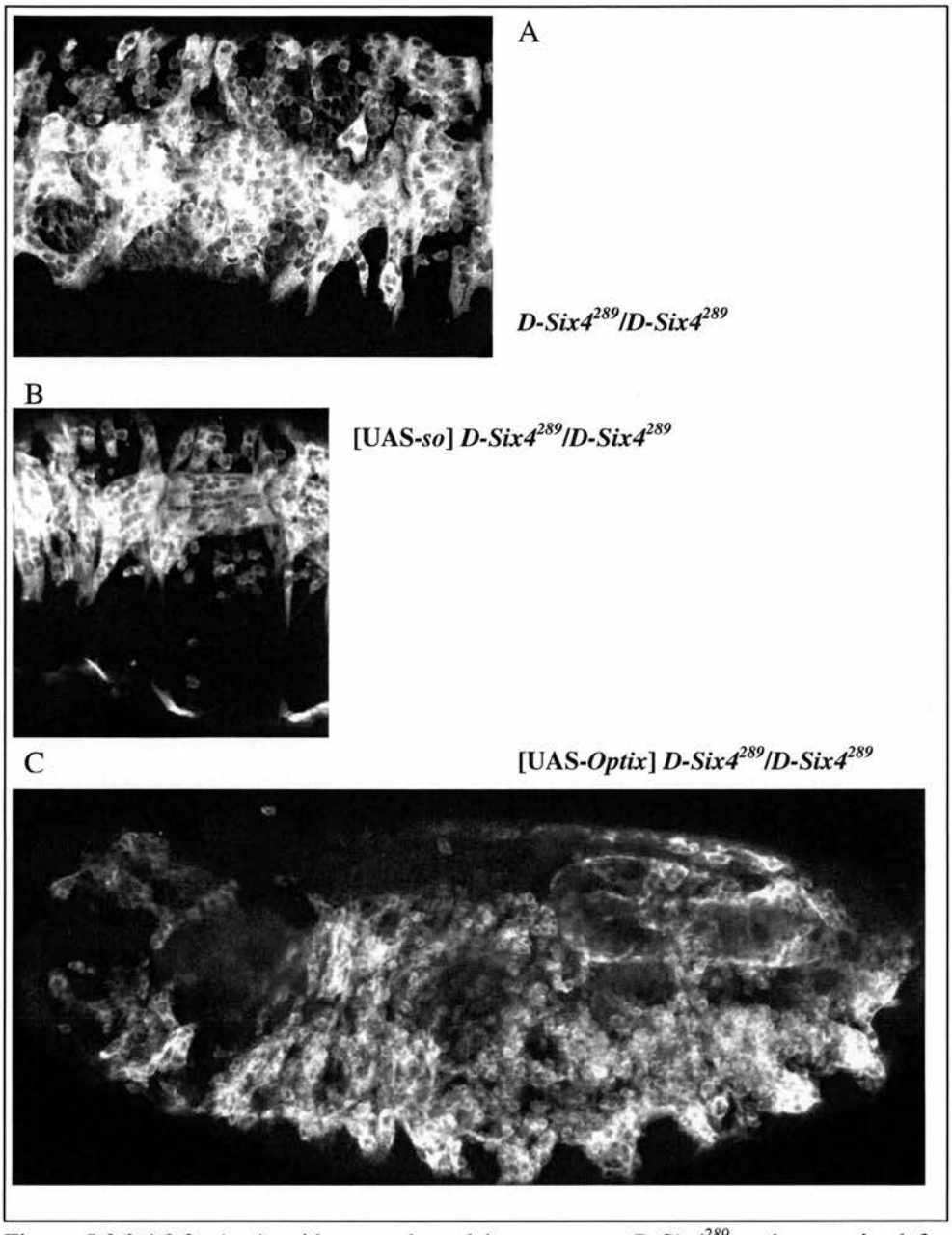


Figure 5.2.3.4.2.3: A: A mid ventro-lateral homozygous *D-Six4*^{289} embryo stained for anti-Myosin displaying a disrupted muscle structure. B: A mid ventro-lateral homozygous *D-Six4*^{289} embryo stained for anti-Myosin expressing p[UAS-*so*] in the mesoderm under the control of the *twist*-GAL4 driver. The phenotype is indistinguishable from the homozygous *D-Six4*^{289} embryo (A) C: A mid ventro-lateral homozygous *D-Six4*^{289} embryo stained for anti-Myosin expressing p[UAS-*Optix*] in the mesoderm under the control of the *twist*-GAL4 driver. Fewer muscles are formed and there are more unfused myocytes.

5.2.3.5 Was there rescue of the gonad phenotype when expressing the *Six* genes?

As expected, the control experiments I carried out all displayed embryos with the previously characterised homozygous *D-Six4*^{289} mutant phenotype, meaning non-coalesced germ cells. As explained previously, the mutant phenotype was previously characterised as the failure of the few SGPs not having undergone cell death to attract germ cells and coalesce into two tightly formed gonads. I redefined these germ cells as scattered. 48% of the phenotypes I observed (n=25) displayed a scattered phenotype and were indistinguishable from the previously characterised mutant phenotype (figure 5.2.3.5.1). Having defined precisely the meaning of scatter, I observed that some embryos displayed what I class as a partial scatter phenotype (figure 5.2.3.5.1). 52% of embryos displayed the partial scatter phenotype. I used this percentage as the baseline percentage. In the experiments described below, a percentage of embryos displaying partial scatter of the germ cells higher than 52% was considered a partial rescue of D-Six4 function.

Expression of D-Six4, So and Optix in the mesoderm of a *D-Six4*^{289} homozygous mutant embryo results in different gonad phenotypes in different proportions.

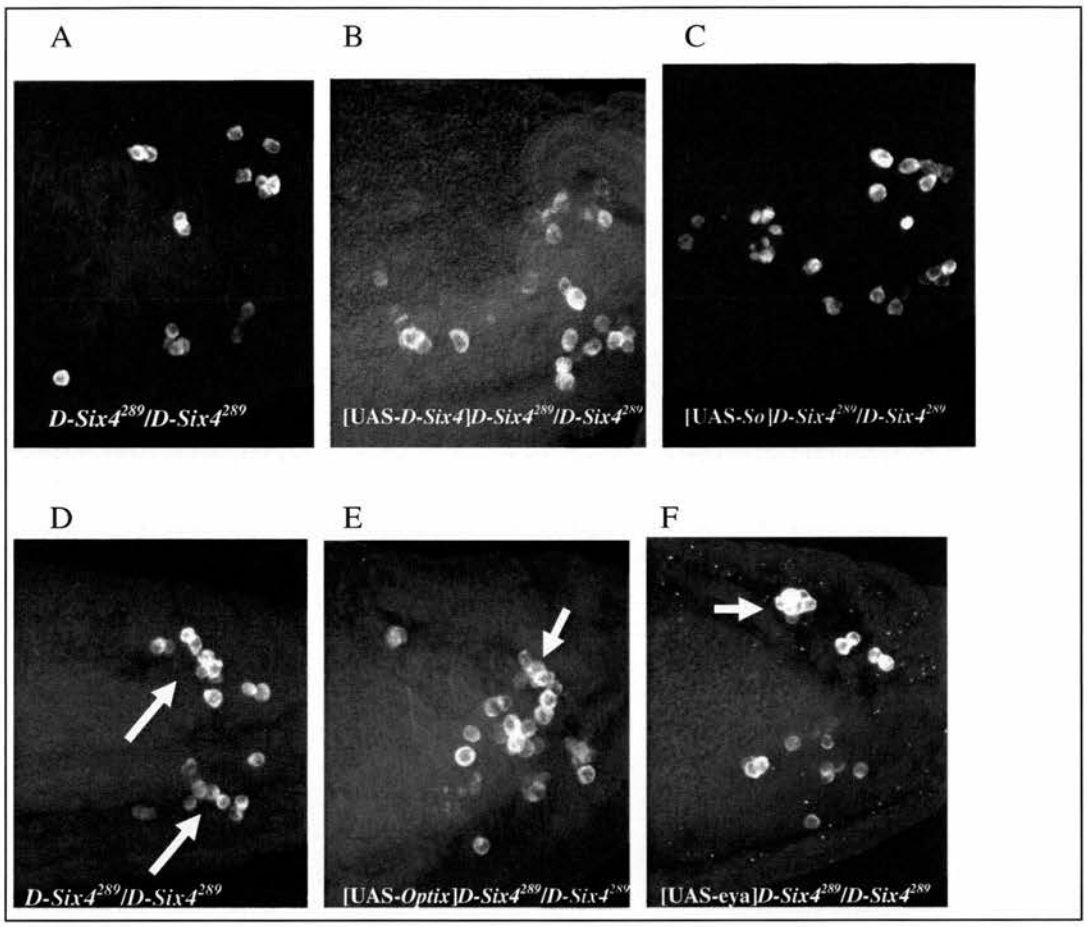


Figure 5.2.3.5.1: Examples of posterior sections of homozygous *D-Six4*^{289} embryos stained for anti-vasa containing one of the p-elements ([UAS-*D-Six4*], [UAS-*so*], [UAS-*Optix*] , [UAS-*eya*]). A, B and C display clusters of no more than 3 germ cells and are therefore classed as scattered. These are z-stacks and although in some cases the germ cells appear clustered this is an artefact of the stacking of different sections. D, E and F display one or more than one cluster of more than 3 germ cells (arrows) and are therefore classed as partially scattered. These p-elements do not affect the mutant phenotype.

5.2.3.5.1 What was the effect of expressing exogenous D-Six4?

Expression of D-Six4 under the control of a *twist*-GAL4 driver results in substantial rescue of the *D-Six4*^{289} homozygous mutant gonad phenotype. I observed embryos displaying tightly coalesced germ cells (figure 5.2.3.5.1.1) forming two compact gonads. These embryos displayed a gonad phenotype indistinguishable from the wild-type phenotype and I therefore classed them as fully coalesced. 78% of the embryos I assessed (n=105) were classed as fully coalesced (figure 5.2.3.5.1.2). The remainder, 22%, of the embryos displayed a variety of phenotypes, all displaying at least one cluster of more than three germ cells (figures 5.2.3.5.1.2 and 5.2.3.5.1.3). I classed these embryos as partially scattered.

78% of *D-Six4*^{289} homozygous mutant embryos display rescue of D-Six4 function when expressing D-Six4 under the control of a *twist*-GAL4 driver. The remainder display partial scattering. Although, the percentage of embryos displaying this phenotype is lower than 52% and therefore is theoretically not indicative of partial rescue of D-Six4 function, the absence of any embryos displaying a scattered phenotype is indicative of at least partial rescue of D-Six4 function.

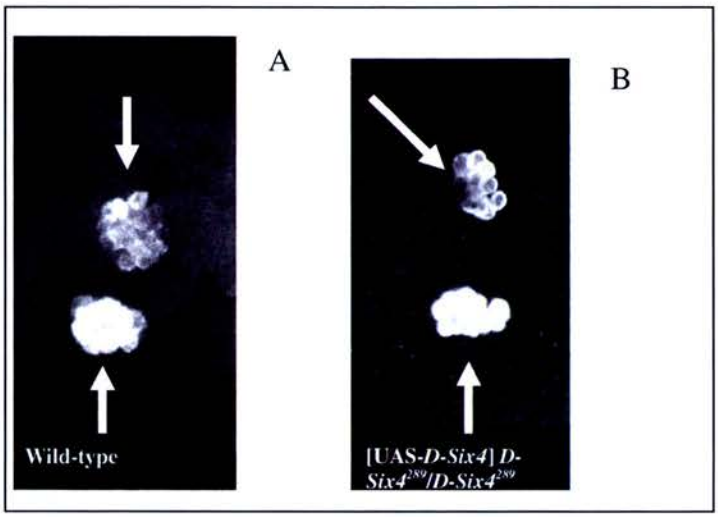


Figure 5.2.3.5.1.1: A: A posterior section of a wild-type embryo stained for anti-vasa showing coalesced germ cells forming two gonads (arrows). B: A posterior section of a homozygous *D-Six4*^{289} embryos stained for anti-vasa expressing [UAS-*D-Six4*] in the mesoderm under the control of a *twist*-GAL4 driver displaying a phenotype indistinguishable from wild type, coalesced germ cells forming two gonads (arrows).

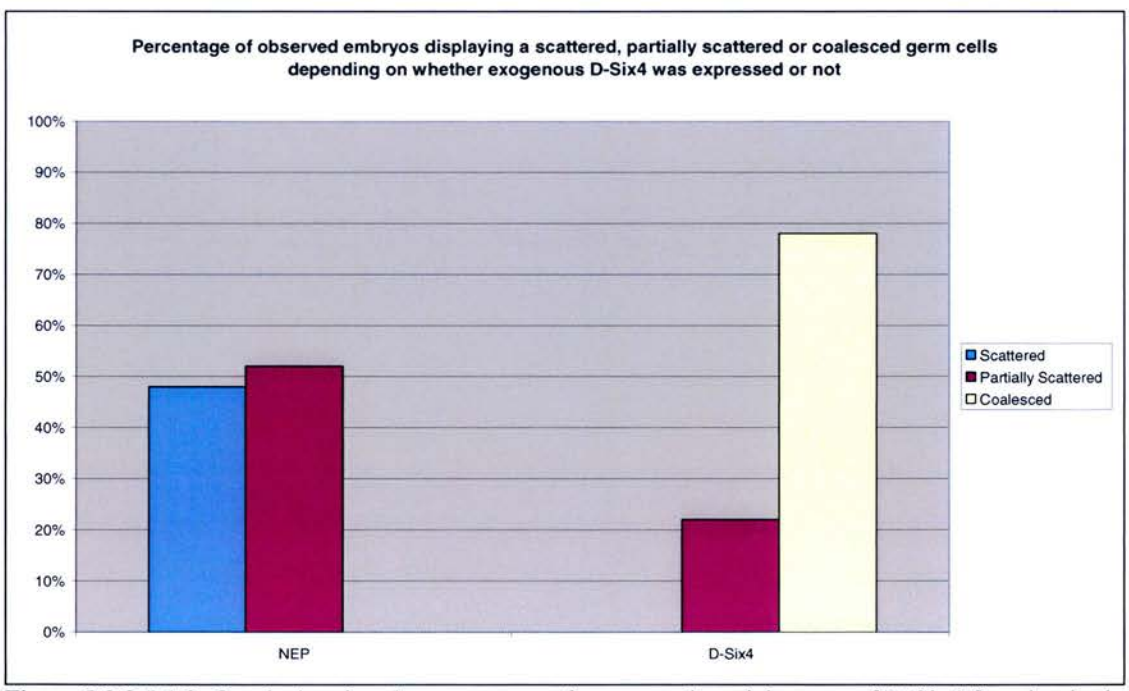


Figure 5.2.3.5.1.2: Graph showing the percentage of rescue and partial rescue of D-Six4 function in the gonad when no exogenous protein (NEP) is expressed compared to when [UAS-*D-Six4*] is being expressed in the mesoderm of a homozygous *D-Six4*^{289} embryo under the control of a *twist*-GAL4 driver (D-Six4).

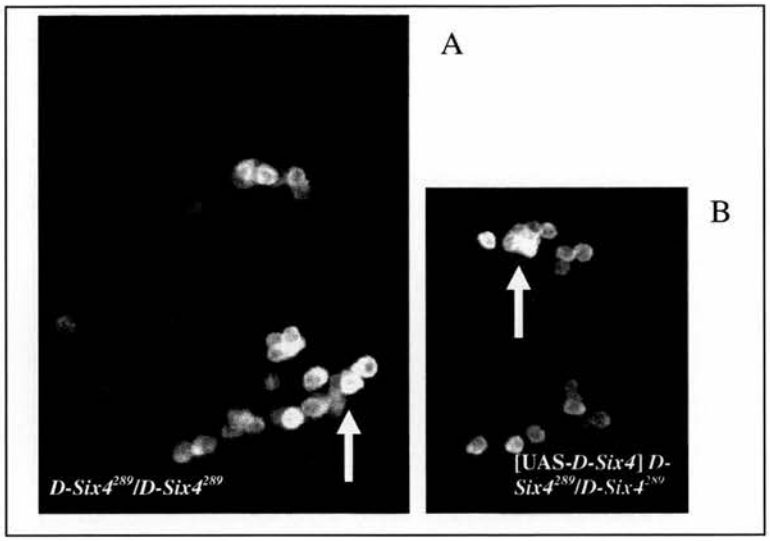


Figure 5.2.3.5.1.3: A: A posterior section of a homozygous *D-Six4*^{289} embryo stained for anti-vasa showing clusters of more than 3 germ cells (arrows) and therefore being classed as partially scattered. B: A posterior section of a homozygous *D-Six4*^{289} embryo stained for anti-vasa expressing [UAS-*D-Six4*] in the mesoderm under the control of a *twist*-GAL4 driver displaying a phenotype indistinguishable from the mutant (A).

5.2.3.5.2 What was the effect of expressing exogenous So?

Expression of So under the control of a *twist*-GAL4 driver indicates partial rescue of the *D-Six4*^{289} homozygous mutant gonad phenotype. I observed embryos displaying a variety of phenotypes, all showing at least one cluster of more than three germ cells (figure 5.2.3.5.2.1). I classed these embryos as partially scattered. 61% of the embryos I assessed (n=141) were classed as partially scattered (figure 5.2.3.5.2.2). Of these a few embryos displayed two larger clusters of germ cells among smaller ones (figure 5.2.3.5.2.3). When I counted the number of germ cells (discussed below), I used these two larger clusters and the numbers I obtained were assumed to be the number of germ cells per gonad in embryos expressing So although they were not gonads *per se*, according to the criteria I previously used. The remainder of the embryos I assessed (39%) showed a phenotype indistinguishable from the previously characterised mutant phenotype (figure 5.2.3.5.2.2). I classed these embryos as scattered.

Partial scatter is observed in 61% of embryos expressing So in a *D-Six4*^{289} homozygous mutant. This is higher than 52% and therefore there is partial rescue of D-Six4 function.

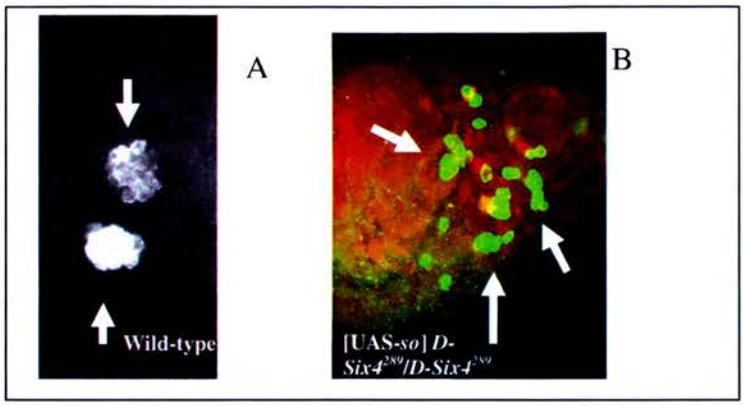


Figure 5.2.3.5.2.1: A: A posterior section of a wild-type embryo stained for anti-vasa displaying coalesced germ cells into two tightly formed gonads (arrows). B: A posterior section of a homozygous *D-Six4*^{289} embryo stained for anti-vasa expressing [UAS-*so*] in the mesoderm under the control of a *twist*-GAL4 driver displaying more than one cluster of more than 3 germ cells (arrows) and therefore was classed as partially scattered.

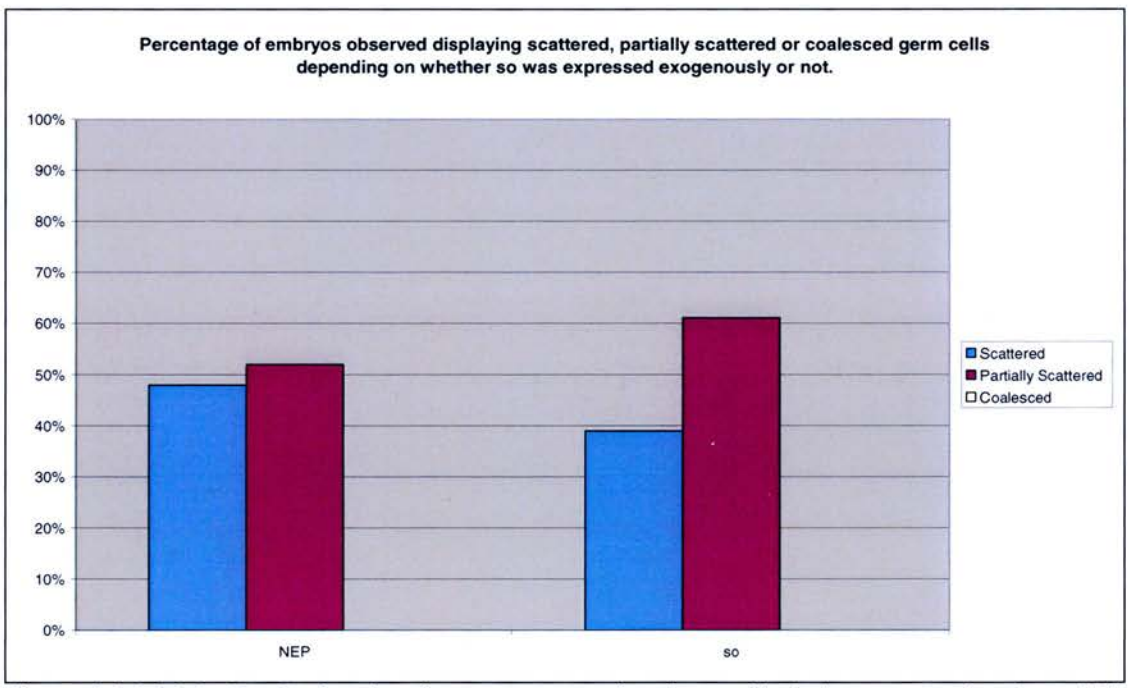


Figure 5.2.3.5.2.2: Graph showing the percentage of embryos displaying scattered and partially scattered germ cells when no exogenous protein, (NEP), is being expressed in the mesoderm of a homozygous *D-Six4*^{289} embryo compared to when [UAS-*so*], (so), is expressed in the mesoderm of a homozygous *D-Six4*^{289} embryo under the control of a *twist*-GAL4 driver.

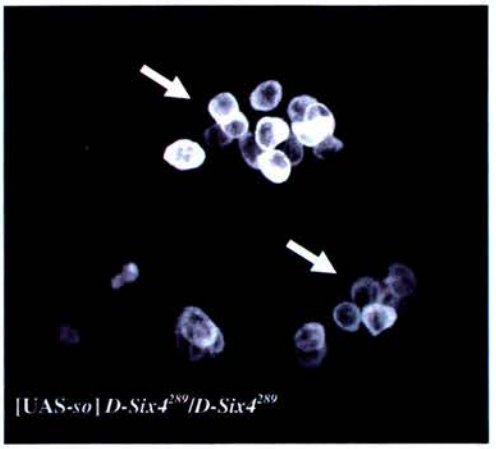


Figure 5.2.3.5.2.3: A posterior section of a homozygous *D-Six4*^{289} embryo stained for anti-vasa expressing [UAS-*so*] in the mesoderm under the control of a *twist*-GAL4 driver displaying two larger clusters of germ cells among smaller ones. These were used in the counting of germ cells.

5.2.3.5.3 What was the effect of expressing exogenous Optix?

Expression of Optix in *D-Six4*^{289} homozygous mutant embryos using a *twist*-GAL4 driver does not rescue the mutant phenotype.

I observed embryos displaying phenotypes showing at least one cluster of more than three germ cells (figure 5.2.3.5.3.1). 52% of the embryos I assessed (n=114) displayed this phenotype (figure 5.2.3.5.3.2). I classed them as partially scattered but since the percentage was the same as in the mutant embryos not expressing the exogenous protein, I considered that there was no rescue of D-Six4 function through the expression of Optix. I observed embryos displaying scattering of germ cells and classed them as scattered (figure 5.2.3.5.3.1). 36% of the embryos assessed were classed as scattered (figure 5.2.3.5.2.2). The remainder embryos, 12%, displayed a phenotype which at first I mistook for wild-type. Indeed, the germ cells were tightly coalesced in two compact gonads (figure 5.2.3.5.3.3). However, they seemed bigger than wild-type gonads (figure 5.2.3.5.3.3). The germ cells' morphology seemed to remain the same indicating that the increase in size was a result of an increase in number as opposed to an increase in the size of the individual germ cells. I tested this by counting the number of germ cells. As I discuss below, the number of germ cells in the gonads present in the embryos expressing Optix under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} mutant phenotype is greater than in a wild-type embryo. These embryos were distinguishable from the wild-type embryos, consequently, I could not class these embryos as fully rescued. In the muscle, expression of Optix under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} mutant embryo results in a proportion of embryos displaying a more severe phenotype which I classed as worsened. The phenotype observed in the germ cells of

these embryos does not appear to be a more severe phenotype than the mutant one. Rather it seems to be a new phenotype. To differentiate between the previously characterised mutant term, I classed these embryos as super-mutant.

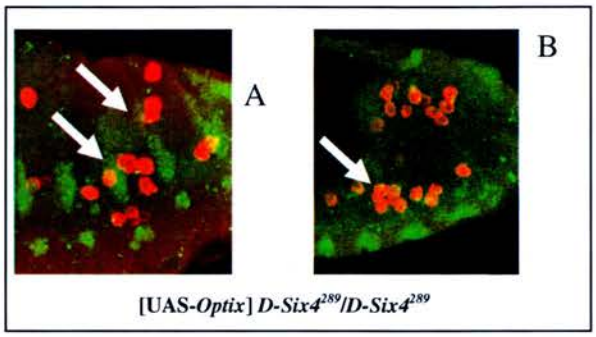


Figure 5.2.3.5.3.1: A: A posterior section of a homozygous *D-Six4*^{289} embryo stained for anti-vasa expressing [UAS-*Optix*] in the mesoderm under the control of a *twist*-GAL4 driver displaying clusters of no more than 3 germ cells (arrows) and therefore being classed as scattered. B: A posterior section of a homozygous *D-Six4*^{289} embryos stained for anti-vasa expressing [UAS-*Optix*] in the mesoderm under the control of a *twist*-GAL4 driver displaying clusters of more than 3 germ cells (arrows) and therefore being classed as partially scattered.

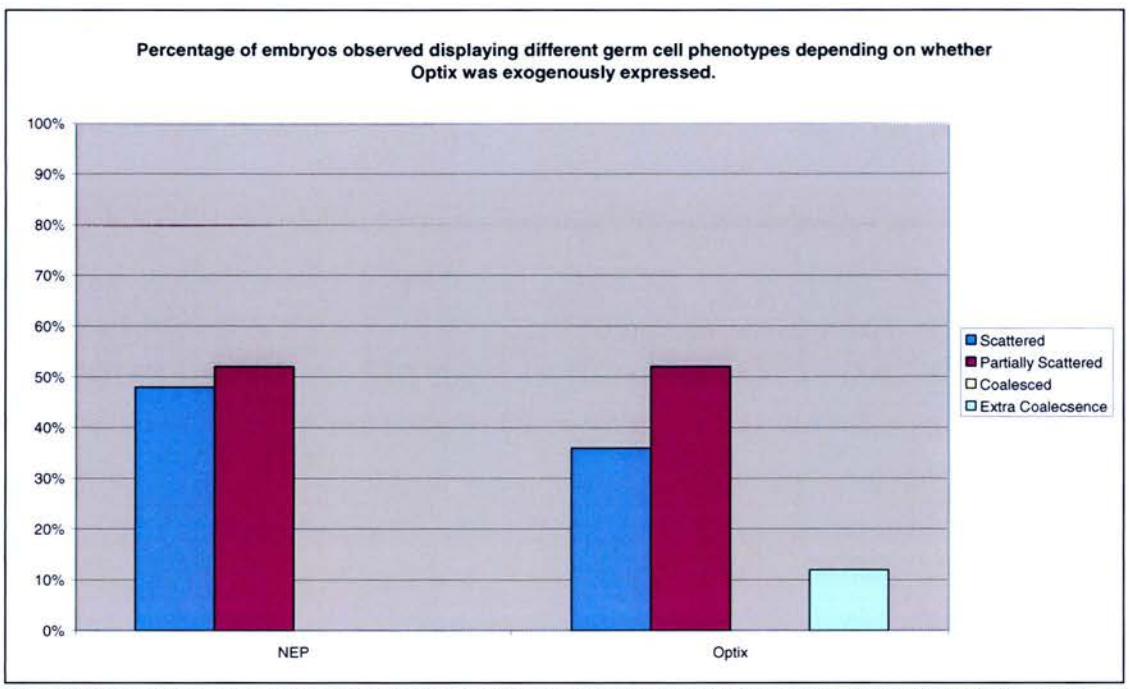


Figure 5.2.3.5.3.2: Graph showing the percentage embryos displaying scattered and partially scattered germ cells when no exogenous protein, (NEP), was expressed in the mesoderm of a homozygous *D-Six4*^{289} embryo compared to when [UAS-*Optix*], (Optix), was expressed in the mesoderm of a homozygous *D-Six4*^{289} embryo under the control of a *twist*-GAL4 driver.

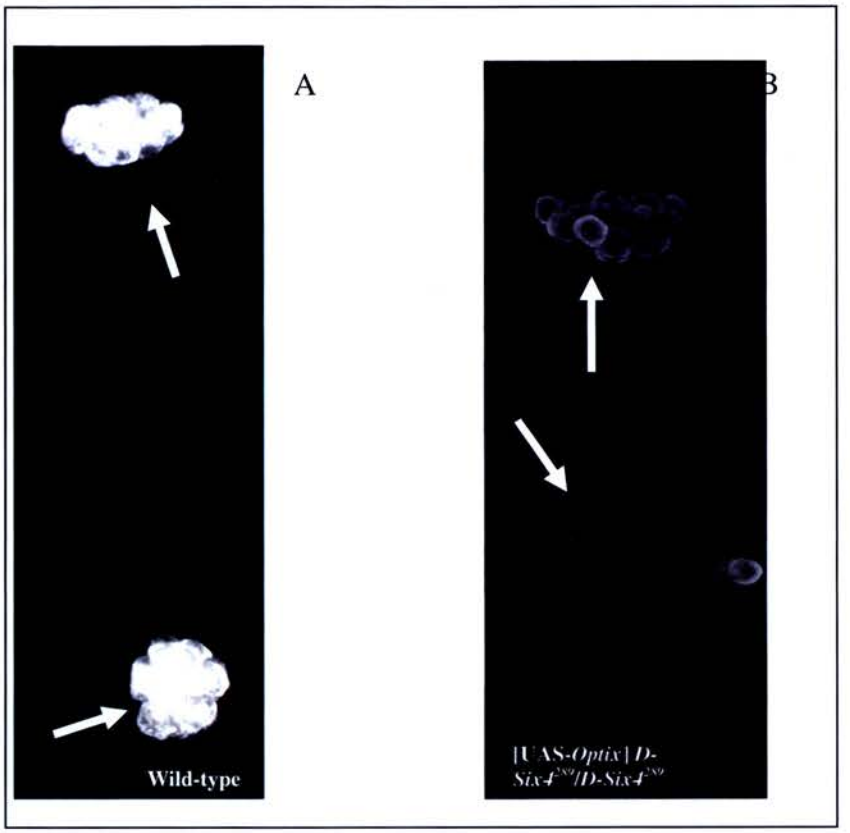


Figure 5.2.3.5.3.3: A: A posterior section of a wild-type embryo stained for anti-vasa showing coalesced germ cells forming two gonads (arrows). B: A posterior section of a homozygous *D-Six4*^{289} embryo expressing [UAS-*Optix*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-vasa and showing two bigger gonads with what appears to be more germ cells (arrows).

5.2.3.5.4 What did the counting of germ cells reveal?

I counted the number of germ cells in the gonads of embryos expressing Optix under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} homozygous mutant and compared the numbers obtained to the ones obtained when counting the germ cells in the gonads of embryos expressing D-Six4 and So respectively under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} homozygous mutant (figure 5.2.3.5.4.1). I scored the gonads in the fully coalesced embryos expressing D-Six4 and in the partially scattered embryos expressing So (explained previously). I also calculated the number of germ cells in the *D-Six4*^{289} mutant embryos not expressing exogenous proteins and in wild-type embryos. As mentioned, the germ cells in the mutant embryos do not coalesce and so I counted the scattered germ cells.

I established that the average number of germ cells in a wild-type gonad is 31 (figure 5.2.3.5.4.1). This number is subject to controversy as indeed it has been reported that there as few as 12 germ cells (Gilboa and Lehmann, 2006) in each embryonic gonad and as many as 31 (Ken Howard, personal communication). Ultimately upon consultation with Dr. Howard, I am confident that my number is correct (Ken Howard, personal communication). It remains to be elucidated why there are these discrepancies. A homozygous *D-Six4*^{289} mutant displays half as many germ cells, 16 (figure 5.2.3.5.4.1). Expressing D-Six4 under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} homozygous mutant re-establishes the wild-type number of germ cells, 30 (figure 5.2.3.5.4.1). Expression of So under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} homozygous mutant gives rise to clusters of up to 20 germ cells (figure 5.2.3.5.4.1). Expression of Optix under the control of a *twist*-

GAL4 driver in a *D-Six4*^{289} homozygous mutant results in slightly bigger gonads containing on average 50 germ cells (figure 5.2.3.5.4.1).

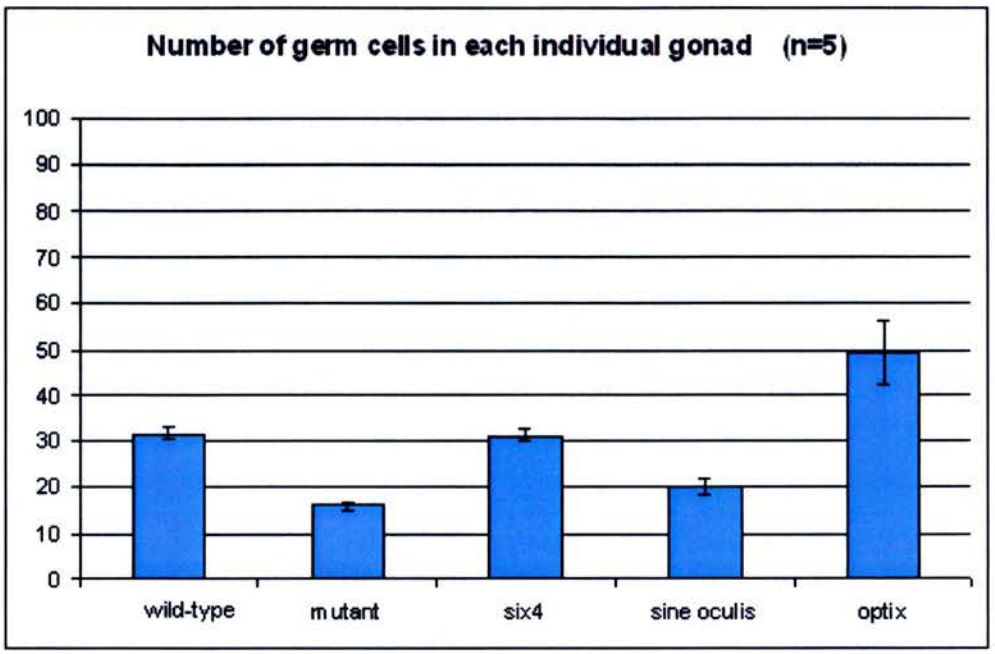


Figure 5.2.3.5.4.1: Number of germ cells observed in individual gonads. Germ cells were counted in 5 gonads of wild-type embryos (wild-type), *D-Six4*^{289} mutant embryos expressing exogenous D-Six4 (six4), *D-Six4*^{289} mutant embryos expressing exogenous So (sine oculis), *D-Six4*^{289} mutant embryos expressing exogenous Optix (optix) (data ± sem, n=5). In *D-Six4*^{289} mutant embryos (mutant) all the germ cells present in the posterior of the embryo were counted and the total number was divided by two to obtain a number of germ cells per gonad. This graph shows that when expressing exogenous D-Six4 in *D-Six4*^{289} mutant embryos it re-establishes the wild-type number of germ cells whereas when expressing exogenous So in *D-Six4*^{289} mutant embryos, the number of germ cells is higher than in the *D-Six4*^{289} mutant but lower than in wild-type. Finally, expressing exogenous Optix in *D-Six4*^{289} mutant embryos results in an increase in the number of germ cells compared to wild-type.

5.2.3.6 Was there rescue of the muscle and gonad phenotype when expressing Eya?

Eya is endogenously expressed in *D-Six4*^{289} homozygous mutant embryos and yet appears to not be able to compensate for the loss of function of D-Six4. Eya and D-Six4 mutants have similar phenotypes in the muscle and the gonad and therefore may carry out similar functions. Alternatively they function together in order to drive the development of these organs. In order to shed light on the relationship between Eya and D-Six4, Eya was overexpressed under the control of a *twist*-GAL4 driver in *D-Six4*^{289} mutant embryos and the muscle and gonad phenotypes were assessed. I observed no rescue of the *D-Six4*^{289} homozygous mutant phenotype when expressing Eya either in the muscle or the gonad.

In the gonad, half of the embryos assessed displayed a phenotype indistinguishable from the previously characterised mutant phenotype (n=25) and the other half displayed a partial scatter phenotype as I described in the *D-Six4*^{289} mutants not expressing exogenous proteins (figure 5.2.3.6.1). Because the percentage of partial scatter in the *D-Six4*^{289} homozygous mutants expressing exogenous Eya was the same as the percentage of partial scatter in the *D-Six4*^{289} homozygous mutant not expressing exogenous proteins, I concluded that expression of Eya did not result in an increase in the number of embryos displaying partial scatter and thus did not result in a partial rescue of D-Six4 function in the gonad. In the muscle, although full rescue as I defined it was not observed, rescue of ventral acute muscle 3 was observed in all the embryos assessed (n=10) (figure 5.2.3.6.2). Although I have not devised criteria defining partial rescue of the mutant phenotype because of the difficulties involved, the consistent presence of this muscle, usually absent in the *D-*

Six4^{289} mutant suggests a partial rescue of D-Six4 function in the development of that muscle. Misexpression of Eya in a wild-type embryo results in a duplication of ventral acute muscle 3. It is therefore relevant to question whether the partial rescue of ventral acute 3 in a *D-Six4*^{289} homozygous mutant expressing Eya is the result of a rescue of the endogenous ventral acute muscle 3 or of the duplicated one. This would additionally assume that the duplicated VA3 only requires the function of Eya to develop.

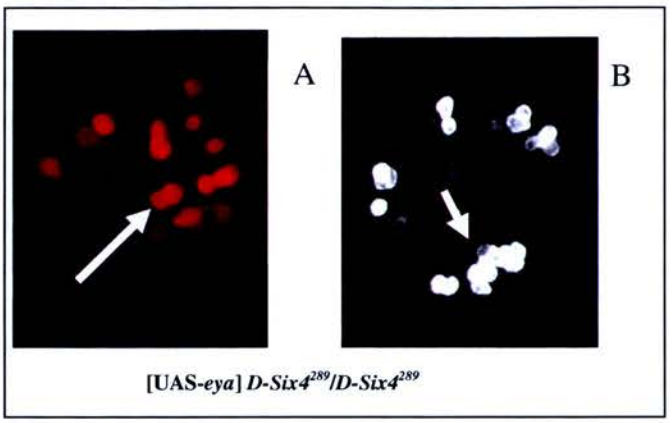


Figure 5.2.3.6.1: A: A posterior section of a homozygous *D-Six4*^{289} embryo expressing [UAS-*eya*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-vasa displaying clusters of no more than three germ cells (arrow) and therefore being classed as scattered. B A posterior section of a homozygous *D-Six4*^{289} embryo expressing [UAS-*eya*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-vasa displaying clusters of more than three germ cells (arrow) and therefore being classed as partially scattered.

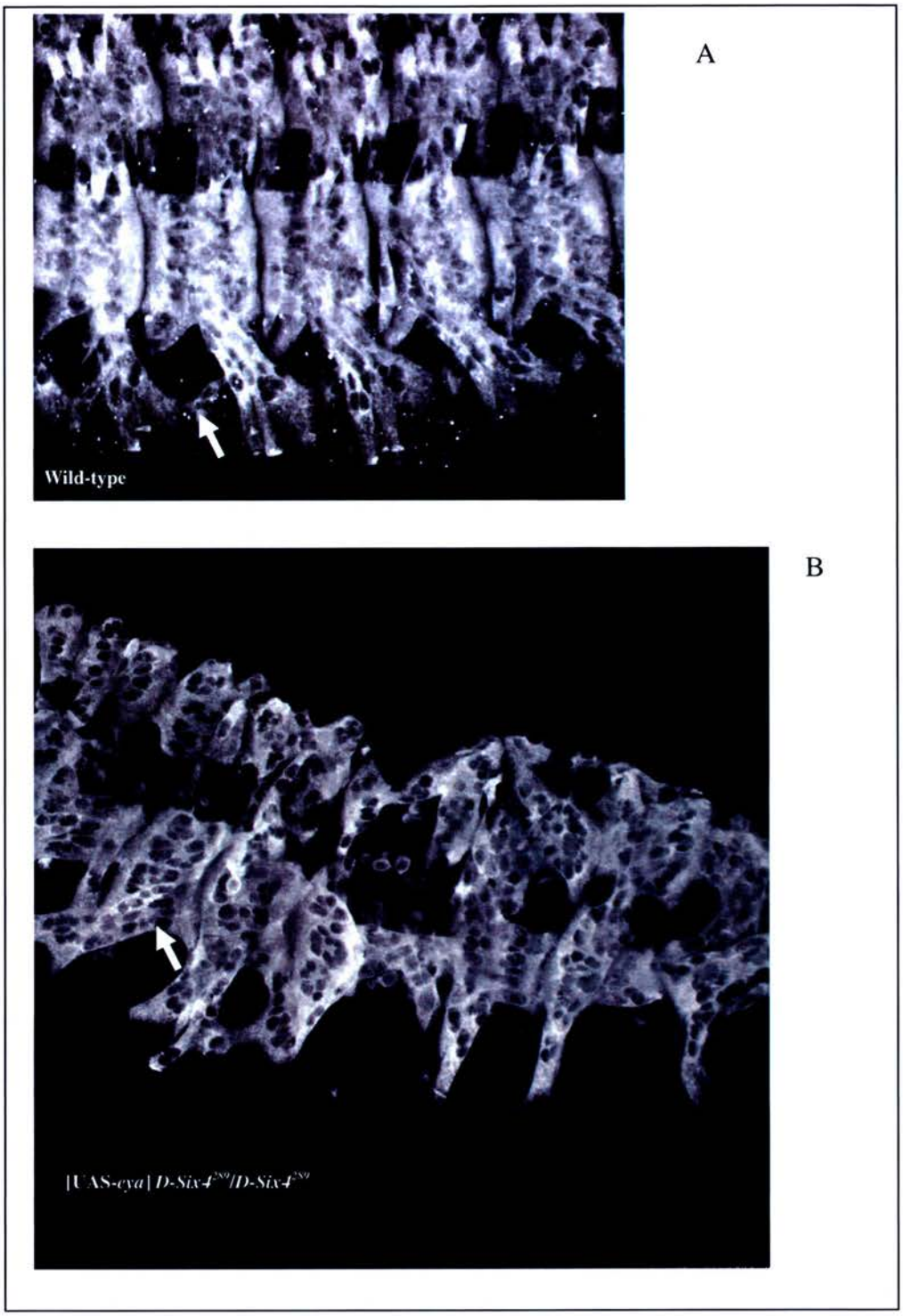
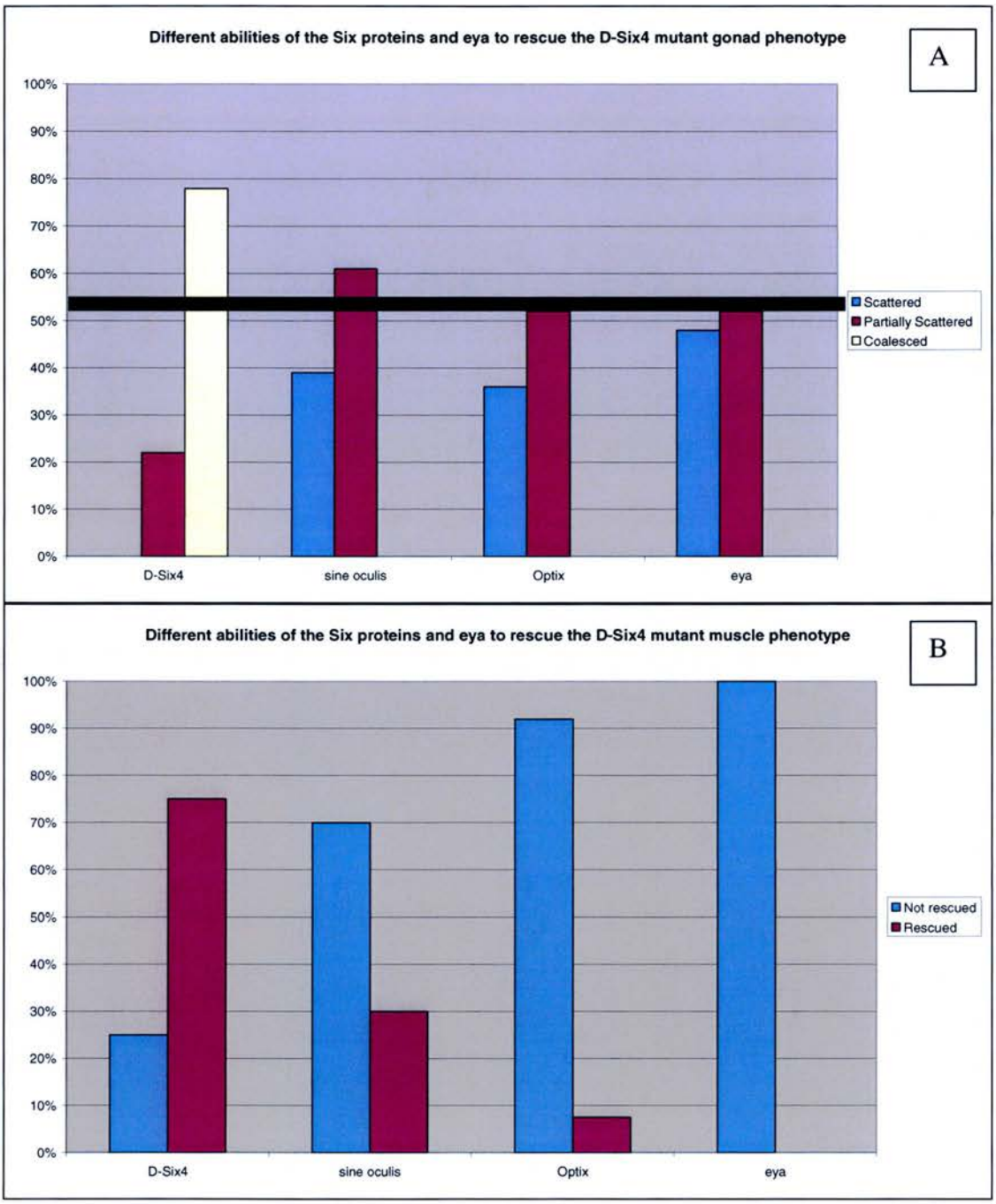


Figure 5.2.3.6.2: A: A mid ventro-lateral wild-type embryo stained for anti-Myosin displaying a regular array of muscles. The arrow points to VA3. B. A ventro-lateral section of a homozygous *D-Six4*^{289} embryo expressing [UAS-*eya*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-Myosin displaying no rescue but duplication of VA3 (arrow).

5.2.3.7 How does rescue of D-Six4 function compare in the muscle and the gonad when expressing these different proteins?

Rescue and partial rescue of D-Six4 function in the gonad is assessed by calculating the proportion of fully coalesced and partially scattered (over 52%) germ cells in the embryos expressing the exogenous proteins in comparison to the ones not expressing exogenous proteins (figure 5.2.3.7.1). Expressing Optix and Eya exogenously under the control of a *twist*-GAL4 driver does not rescue D-Six4 function in the gonad (figure 5.2.3.7.1). I observed partial rescue of D-Six4 function when expressing So and full rescue when expressing D-Six4 (figure 5.2.3.7.1). In the muscle, expressing Eya exogenously under the control of a *twist*-GAL4 driver does not rescue D-Six4 function (figure 5.2.3.7.1). I observed rescue of D-Six4 function when expressing So, D-Six4 and Optix (figure 5.2.3.7.1). These different abilities of these proteins to rescue some aspects of the mutant phenotype and not others indicate different functions of D-Six4 in these processes. This will be discussed further in the discussion.


Different abilities of the Six proteins and eya to rescue the D-Six4 mutant gonad phenotype
A
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
D-Six4
sine oculis
Optix
eya
Scattered
Partially Scattered
Coalesced
Different abilities of the Six proteins and eya to rescue the D-Six4 mutant muscle phenotype
B
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
D-Six4
sine oculis
Optix
eya
Not rescued
Rescued

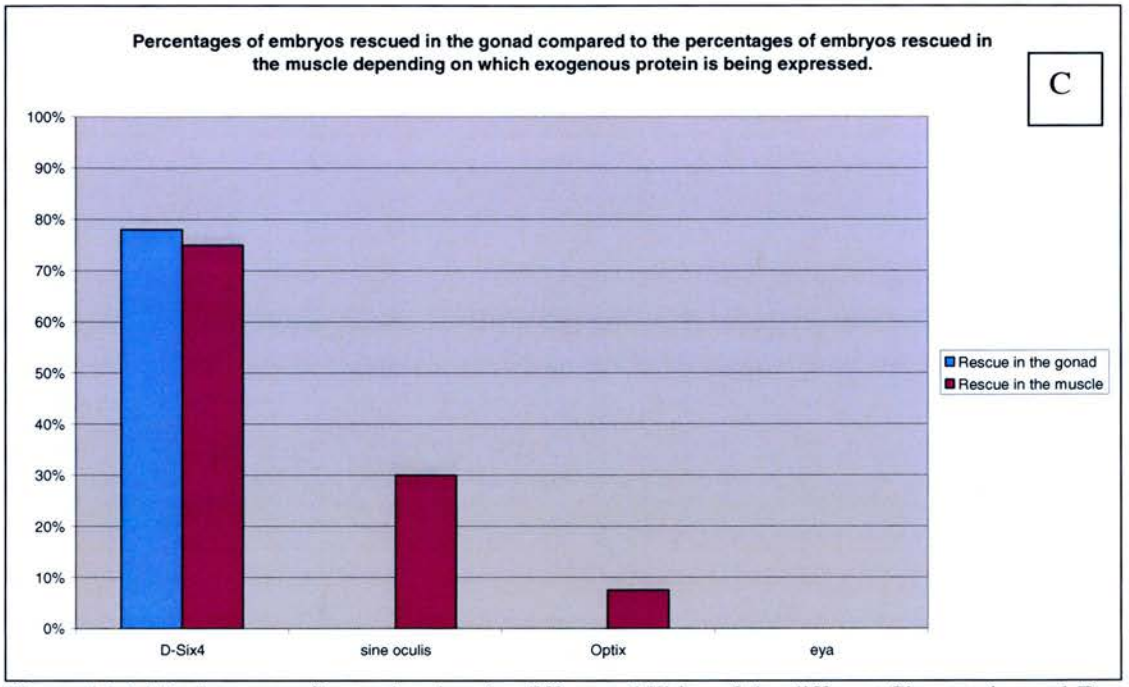


Figure 5.2.3.7.1: Summary figure showing the different abilities of the different Six proteins and Eya to rescue the mutant gonad phenotype (A), the mutant muscle phenotype (B) and their different abilities in each tissue (C). A: The horizontal black line shows the baseline of 52% of partial scatter observed in the mutant. Percentages of partial scatter greater than 52% (above the line) indicate partial rescue of D-Six4 function. Partial rescue of D-Six4 function is observed when expressing So exogenously. Total rescue of the gonad phenotype (coalescence of the germ cells) is only achieved when expressing D-Six4 exogenously. B: Total rescue of the muscle phenotype is achieved when expressing D-Six4, So and Optix exogenously but in different percentages. No rescue is observed when expressing Eya exogenously. C: The ability of each Six protein and Eya to rescue D-Six4 function in the gonad (to the extent of total coalescence of germ cells) and the muscle is shown.

5.2.3.8 Did misexpression of these proteins in the mesoderm have any effect?

In the progeny of the crosses analysed above, the expression of D-Six4, So, Optix and Eya, respectively, was driven under the control of a *twist*-GAL4 driver in the mesoderm of half of the heterozygous embryos (figure 5.2.3.8.1). I analysed these embryos to examine the effect of misexpression of these proteins on heterozygous (wild-type) muscle and gonad development.

At the embryonic level, the staining procedures I carried out resulted in the homozygous *D-Six4*^{289} mutants not staining for beta-galactosidase and the three other genotypes staining for beta-galactosidase (figure 5.2.3.8.1). Of these three genotypes, only one would potentially be expressing the exogenous proteins under the control of a *twist*-GAL4 driver (figure 5.2.3.8.1). I was able to differentiate the balancer homozygotes from the two other genotypes. Indeed, balancer homozygotes are easily recognisable as embryos appear distorted morphologically. In the muscle, very few muscles stain for Myosin and the general phenotype is severely mutant and distinguishable from the *D-Six4*^{289} mutant phenotype. In the gonad, hardly any germ cells are present. This is consistent with expectations as indeed balancers were designed to be homozygous lethal. I was not able to distinguish between the two other genotypes and assumed that they were represented in equal amounts meaning that half of the embryos had the genotype expressing the exogenous proteins in the mesoderm and the other half had the genotype not expressing these exogenous proteins (figure 5.2.3.8.1).

D-Six4^{289} heterozygotes have a mutated copy of *D-Six4*^{289} and express one copy of functional endogenous D-Six4. *D-Six4*^{289} is a recessive mutation, therefore

D-Six4^{289} heterozygotes display a wild-type phenotype of the muscle and gonad. Consequently, I classified embryos displaying an indistinguishable phenotype from the previously characterised heterozygotes as wild-type. I classified embryos displaying a different phenotype, distinguishable from the wild-type phenotype as affected. I expected the embryos staining for beta-galactosidase and not displaying the balancer homozygous phenotype to display a wild-type phenotype as indeed in half of them no expression of the exogenous proteins was being driven. Additionally, since in previous controls, the presence of the p-elements had not affected the *D-Six4*^{289} homozygous mutant phenotype, I did not expect it to affect the heterozygous mutant one either. Because I assumed that the two genotypes were represented in equal amounts, the percentage of affected embryos observed was in relation to both genotypes. Indeed, 50% or less of affected embryos suggests that they had the genotype expressing the exogenous proteins, more, would suggest that they were not present in equal amounts. Misexpression of each of these four proteins is discussed in turn.

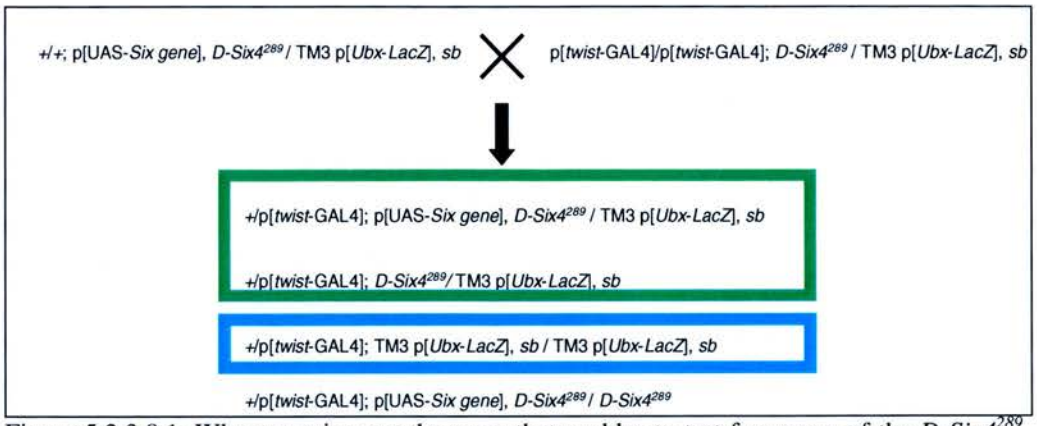


Figure 5.2.3.8.1: When carrying out the cross that enables to test for rescue of the *D-Six4*^{289} mutant phenotype, four genotypes arise. The genotype in the blue square is the homozygous balancer which was not used in the comparisons. The two genotypes in the green square are the heterozygous genotypes containing one wild-type copy of D-Six4 and one mutated one.

5.2.3.8.1 What was the effect of misexpressing So?

Misexpressing So in heterozygous *D-Six4*^{289} mutants under the control of a *twist*-GAL4 driver results in there being two expressed wild-type copies of endogenous Eya, one expressed wild-type copy of endogenous D-Six4 and one expressed wild-type copy of exogenous So (table 5.2.3.8.1.1). I observed no discernible phenotypical effect on the muscle structure in the *D-Six4*^{289} heterozygous mutants which appeared indistinguishable from wild-type (figure 5.2.3.8.1.1). I classed all the embryos observed as wild-type (n=50). Either misexpression of So resulted in death of all the embryos expressing So before egg-laying or the misexpression of So did not have a phenotypic effect. In view of the fact that a gonad mutant phenotype was observed, I proposed that the *D-Six4*^{289} heterozygous mutant embryos expressing So under the control of a *twist*-GAL4 driver did not die before egg laying and therefore misexpression of So does not have a discernible mutant muscle phenotype.

The gonad however is disrupted in half of the heterozygous embryos (52%) (n=631). Indeed, I observed an additional cluster of five to ten germ cells anterior to the two wild-type gonads (figure 5.2.3.8.1.2). These germ cells seem to have remained at the position at which they were after exiting the gut after germ band extension and not to have migrated along the dorsal side of the embryo following germ band retraction. I classed these embryos as affected. Half of the heterozygotes displayed this phenotype and if the assumption that the two genotypes are present in equal amounts is verified, then this means that misexpression of So in the mesoderm of *D-Six4*^{289} heterozygous mutant embryos under the control of a *twist*-GAL4 driver results in disruption of the gonad phenotype in all embryos assessed.

Previously, the control experiments I had carried out had been in order to ensure that the presence of the p-elements had no effect on the *D-Six4*^{289} mutant phenotype. Because I showed that they didn't, I assumed that they did not have an effect on the heterozygous phenotype. In view of the result I obtained in the heterozygous mutant expressing So under the control of a *twist*-GAL4 driver, I tested out this assumption. I thereby wanted to verify that the observed phenotype was the result of So being expressed under the control of a *twist*-GAL4 driver and not simply the result of the presence of the [UAS-*so*] p-element. Unexpectedly, control crosses revealed that embryos with the genotype p[UAS-*so*], *D-Six4*^{289} / TM3 p[*Ubx*-LacZ], *sb* displayed the same phenotype as the one observed in the embryos with the genotype [*twist*-GAL4]/+ ; p[UAS-*so*], *D-Six4*^{289} / TM3 p[*Ubx*-LacZ], *sb*. In view of this, I hypothesised that the insertion of the p[UAS-*so*] element was near a promoter driving basal expression of So even in the absence of the p[*twist*-GAL4] element. This means the assumption that the two heterozygous genotypes are present in equal amounts just because 50% or less show a mutant phenotype is wrong. Indeed, the embryos not having the *twist*-GAL4 driver are nonetheless expressing So at a basal level. These two genotypes may be present in equal amounts or not. These two genotypes are indistinguishable. Therefore it is not possible to assess whether a genotype is more present than the other. All that can be said is that of the embryos present, irrespective of their genotype, a proportion is expressing So. In order to explain the lack of phenotype of this putative basal expression of So in the *D-Six4*^{289} homozygous mutant and a discernible phenotypic effect in the *D-Six4*^{289} heterozygous mutant, I suggested that the observed phenotypic effect in the

heterozygote was the result of a competition between So and D-Six4 for their putative common cofactor Eya.

Expressed protein / Gene	E D-Six4	X D-Six4	E So	X So	E Optix	X Optix	E Eya	X Eya
D-Six4	1	1					2	
so	1			1			2	
Optix	1					1	2	
eya	1						2	1

Table 5.2.3.8.1.1: Endogenous (E) and exogenous (X) expressed protein copies of D-Six4, So, Optix and Eya (in green) when misexpressing under the control of a *twist*-GAL4 driver *D-Six4*, *so*, *Optix* and *eya* (in purple) in the mesoderm of a heterozygous *D-Six4*^{289} embryos.

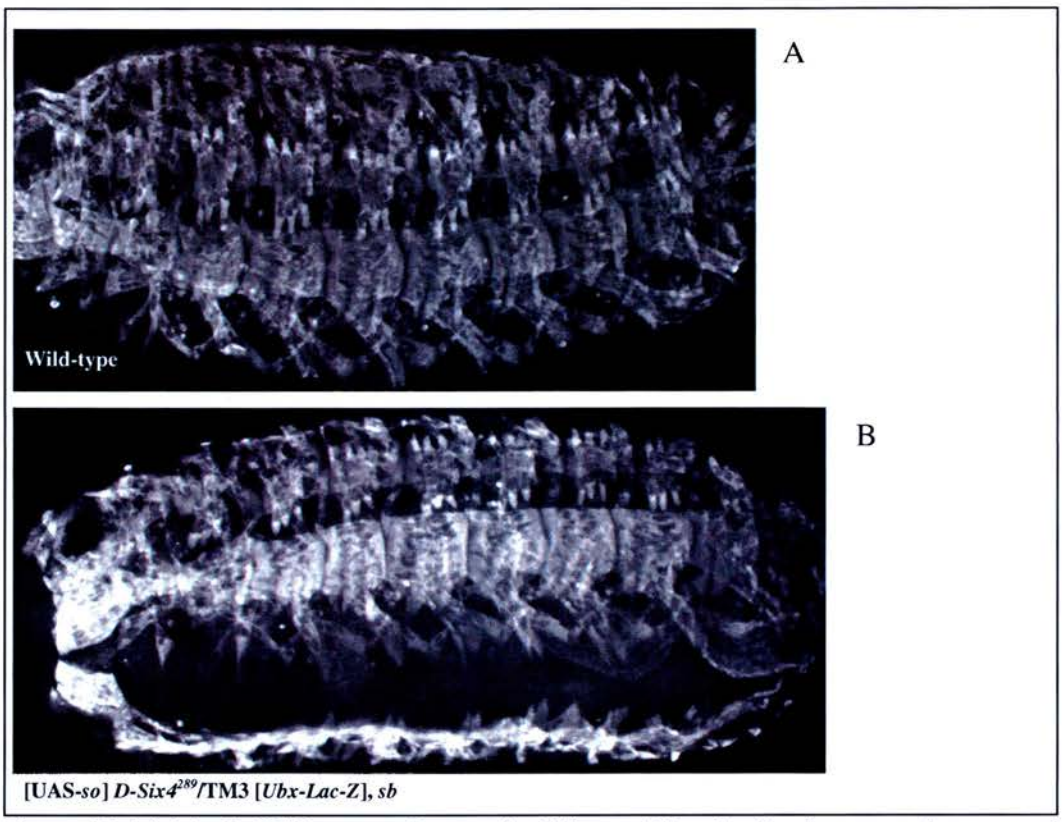


Figure 5.2.3.8.1.1: A: Wild-type embryo stained for anti-Myosin showing a regular array of ventro-lateral muscles. B: A heterozygous *D-Six4*^{289} embryo expressing [UAS-*so*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-Myosin showing a phenotype indistinguishable from wild-type.

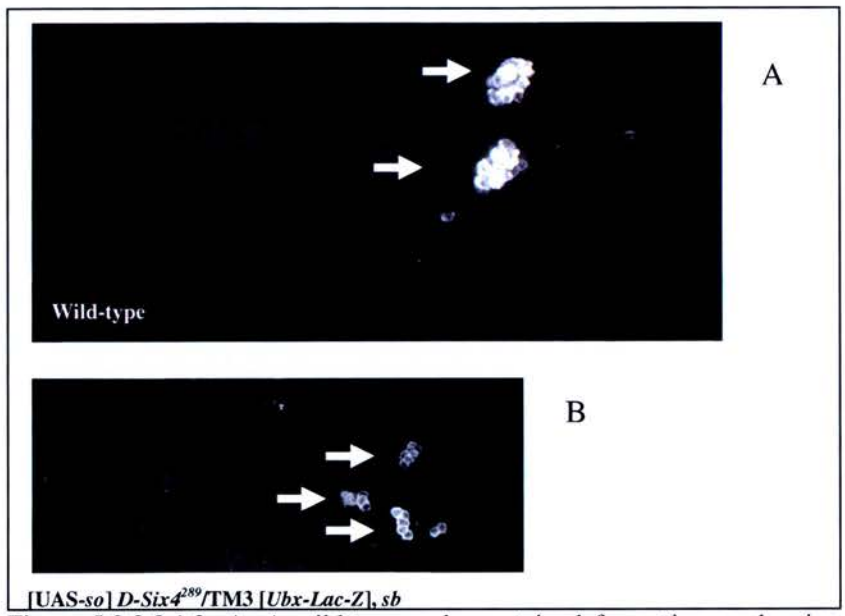


Figure 5.2.3.8.1.2: A: A wild-type embryo stained for anti-vasa showing two tightly coalesced gonads (arrows). B: Heterozygous *D-Six4*^{289} embryo expressing [UAS-*so*] in the mesoderm under the control of a *twist*-GAL4 driver stained for anti-vasa displaying three distinct clusters of germ cells (arrows).

5.2.3.8.2 Do So and D-Six4 compete together for their common cofactor Eya?

I observed that presence of a [UAS-*so*] p-element in a *D-Six4*^{289} mutant embryo result in an affected phenotype in the embryos expressing an endogenous copy of D-Six4 (the heterozygotes) but has no effect on the homozygous mutant. The phenotype observed is the disruption of the patterning of the mesoderm through the failure of some germ cells to coalesce into the gonad. I propose that these results suggest that exogenous So expression may interfere with endogenous D-Six4 function. A possible mechanism for this would be that So competes with D-Six4 for their putative common cofactor Eya. I hypothesised that this competition might be tested by altering relative gene dosage. One prediction would be that expression of two copies of the endogenous D-Six4 protein would outcompete the singly expressed exogenous copy of the So protein and would rescue the mutant phenotype observed in the heterozygotes.

5.2.3.8.3 Is this competition subject to gene dosage?

I analysed wild-type embryos containing one exogenous copy of p[UAS-*so*] and showed that fewer embryos displayed a disrupted gonad phenotype than was observed when expressing one exogenous copy of p[UAS-*so*] in a *D-Six4*^{289} heterozygous mutant (14% as opposed to 52%) (n=50) (table 5.2.3.8.3.1). This supported the hypothesis that this competition is subject to gene dosage. I further tested my hypothesis by analysing embryos expressing two copies of endogenous D-Six4 and two copies of exogenous So. I also analysed embryos expressing one copy of endogenous D-Six4 and two copies of exogenous So. In the first case, I expected the gonad phenotype to be affected as indeed the gene dosage would be the same as

when one endogenous copy of D-Six4 and one exogenous copy of So are being expressed. In the second case, I expected the gonad phenotype to be mutant and perhaps even more disrupted as a result of the inbalance of 2:1 copies of exogenous: endogenous So:D-Six4. Contrarily to expectations, both genotypes displayed 100% wild-type phenotypes (n=50 in each case) (table 5.2.3.8.3.1). These results were not supportive of the starting assumptions and so I rejected the gene dosage-related competition hypothesis.

Wild-type gene copies	Wild-Type phenotype	Affected phenotype
2E D-Six4	*100%*	*0%*
1E D-Six4	*100%*	*0%*
2E D-Six4 + 2X So	*100%*	*0%*
1E D-Six4 + 1X So	*48%*	*52%*
2E D-Six4 + 1X So	*86%*	*14%*
1E D-Six4 + 2X So	*100%*	*0%*

Table 5.2.3.8.3.1: Number of wild-type endogenous (E) and exogenous (X) gene copies of D-Six4 and So respectively and the percentage of affected phenotypes observed in the respective embryos.

5.2.3.8.4 Is the third balancer affecting the phenotype?

As an alternative to the above hypothesis, I considered whether the TM3 balancer was affecting the phenotype. Indeed, the affected phenotypes I observed were in all embryos displaying a genotype including both one copy of p[UAS-*so*] and one copy of TM3 p[*Ubx*-LacZ], *sb*. Consequently I hypothesised that the interaction between both might be the cause for this disrupted phenotype. In order to determine whether it was the interaction with the TM3 balancer specifically or whether it was the interaction with any balancer that caused this phenotype, I analysed embryos with the genotype p[UAS-*so*]/ TM6 UBX. All of these displayed a wild type phenotype. I concluded that the phenotype stemmed from the presence of both the TM3 balancer and the [UAS-*so*] p-element.

5.2.3.8.5 What was the effect of misexpressing Optix, D-Six4 and Eya?

Misexpressing Optix in heterozygous *D-Six4*^{289} mutants under the control of a *twist*-GAL4 driver results in there being two expressed wild-type copies of endogenous Eya, one expressed wild-type copy of endogenous D-Six4 and one expressed wild-type copy of exogenous Optix (table 5.2.3.8.1.1). I only observed embryos displaying a phenotype indistinguishable from wild-type embryos. I concluded that this misexpression had no discernible effect either in the muscle or in the gonad.

Misexpressing D-Six4 in heterozygous *D-Six4*^{289} mutants under the control of a *twist*-GAL4 driver results in there being two expressed wild-type copies of endogenous Eya, one expressed wild-type copy of endogenous D-Six4 and one expressed wild-type copy of exogenous D-Six4 (table 5.2.3.8.1.1). Again, no discernible mutant phenotype was observed either in the muscle or in the gonad. I

concluded that overexpressing D-Six4 in the mesoderm has no effect in either the muscle or the gonad.

Misexpressing Eya in heterozygous *D-Six4*^{289} mutants under the control of a *twist*-GAL4 driver results in there being two expressed wild-type copies of endogenous Eya, one expressed wild-type copy of exogenous Eya and one expressed wild-type copy of endogenous D-Six4 (table 5.2.3.8.1.1). I observed no discernible phenotypic effect in the gonads but in the muscle, I observed the duplication of ventral acute muscle 3. Consistent with previous results, I concluded that overexpression of Eya results in the duplication of VA3.

5.3 Discussion

5.3.1 What results did I obtain?

I showed that expressing *Drosophila Six* genes and *eya* under the control of a *twist*-GAL4 driver in the mesoderm of *D-Six4*^{289} homozygous mutant resulted in different levels of rescue of D-Six4 function and in different proportions depending on the protein and the tissue. Expressing exogenous D-Six4 rescues D-Six4 function in the muscle of 75% of embryos assessed and in the gonad in 78%. The remainder of the embryos are not rescued in the muscle and partially rescued in the gonad. Expressing exogenous D-Six4 also re-establishes the wild-type number of germ cells populating the gonad in the fully rescued embryos. Expressing So partially rescues D-Six4 function in the gonad in the 61% of the embryos assessed and rescues D-Six4 function in the muscle in 30%. Expressing exogenous So partially re-establishes the wild-type number of germ cells populating the gonad in the partially rescued embryos. The remainder display a non rescued phenotype both in the muscle and the gonad. Expressing Optix does not rescue D-Six4 function in the gonad but rescues D-Six4 function in the muscle in 8% of the embryos assessed. The remainder display a non rescued phenotype which is either a worsened phenotype (in the muscle) or a super mutant phenotype (in the gonad). Expressing exogenous Optix results in an increase in the number of germ cells populating the gonad compared to wild-type number in the super-mutant embryos. Expressing Eya does not rescue D-Six4 function in the gonad or in the muscle, it only rescues D-Six4 function in the development of VA3.

I also showed that misexpression of D-Six4 and Optix in the mesoderm of *D-Six4*^{289} heterozygous mutants under the control of a *twist*-GAL4 driver had no effect

whereas misexpression of So resulted in a mutant phenotype in the gonad and misexpression of Eya resulted in the duplication of VA3 in the muscle. These results raise questions which I now discuss in turn.

5.3.2 What are the conserved/diverged functions of the Six proteins and Eya in the muscle?

5.3.2.1 What do the rescued embryos suggest?

I observed embryos displaying phenotypes of the ventral and lateral muscles indistinguishable from wild-type in *D-Six4*^{289} homozygous mutants when expressing under the control of a *twist*-GAL4 driver exogenous D-Six4, So and Optix respectively. All three Six proteins show some ability to fully rescue D-Six4 function in the ventral and lateral muscles of *D-Six4*^{289} homozygous mutants. This indicates a potential for D-Six4, So and Optix to carry out similar functions involved in the patterning of the mesoderm and more specifically in the development of the ventral and lateral muscles. As I discussed in chapter 1, endogenous D-Six4 is a transcription factor involved both in protein-protein and protein-DNA interactions. Neither its downstream nor its regulating targets have so far been specifically identified. However candidate cofactors and downstream targets are the other transcription factors expressed at the same time and in the same place as D-Six4. These include *Mef2*, *tinman*, *twist* and *eya* to name but a few (chapter 1). Eya in particular is thought to function together with D-Six4. Evidence for this is the similar mutant phenotypes of single null Eya and single null D-Six4 mutants. Also neither ectopic expression of Eya nor ectopic expression of D-Six4 have a discernible mesoderm phenotype (except for the VA3 muscle). So and Eya are cofactors in the development

of the eye. Whether So and Eya function together to rescue D-Six4 function is unclear. It is unclear whether So and Optix are able to carry out similar protein-protein or protein-DNA interactions or whether they function differently but nonetheless they show some ability to drive the development of the muscle and the gonad. Protein-protein interactions can be tested by histochemically staining for the putative interactors to verify whether they are still expressed in a *D-Six4*^{289} mutant and if so whether their expression patterns coincide with that of exogenous So and Optix. Pull down experiments would test *in vitro* whether these proteins are able to interact. Protein-DNA interactions can also be tested by pull down experiments.

The percentage of embryos showing a wild-type phenotype differs depending on the Six protein expressed. Whereas expressing D-Six4 results in rescue of D-Six4 function in 75% embryos assessed, So and Optix show lower percentages, 30% and 7.5% respectively. I suggest that the variation of phenotypes I observed in the embryos having the same genotype is a consequence of the use of the GAL4/UAS system. Indeed, I suggest that the amount of GAL4 produced differs in the different embryos resulting in different amounts of D-Six4, So and Optix protein being produced. I suggest that the ability of these exogenous proteins to rescue D-Six4 function is dose dependent. I suggest therefore that the rescued phenotypes observed are the result of these exogenous proteins being expressed in amounts that enable them to interact with upstream and downstream targets respectively and drive the development of the muscle. This could be verified by increasing the amount of protein produced and assessing whether this has any effect on the phenotypes observed. This can be done by changing the temperature at which the crosses are carried out. Indeed, expression of effector genes is greater at higher culture

temperatures. Also increasing the number of GAL4 binding sites could increase the expression of effector genes, thus resulting in an increased amount of protein.

Overall, these results suggest that with regards to the functions carried out by D-Six4 in the development of the muscle neither So nor Optix have diverged to the extent of not being able to carry out functions enabling the development of the muscle.

5.3.2.2 What do the mutant embryos suggest?

I also observed embryos displaying phenotypes of the ventral and lateral muscles indistinguishable from mutant in *D-Six4*^{289} homozygous mutants when expressing under the control of a *twist*-GAL4 driver exogenous D-Six4, So and Optix respectively. I also observed these phenotypes when expressing exogenous Eya. In the case of the Six proteins, this indicates an inability of these proteins to rescue D-Six4 function in these embryos. In order to explain these different phenotypes, I have suggested that the inability of these exogenous Six proteins to rescue aspects of the *D-Six4*^{289} homozygous mutant is dose dependent. In these embryos, the amount of D-Six4, So and Optix protein produced does not enable these proteins to interact with upstream and/or downstream targets to drive the development of the muscle and rescue D-Six4 function. In the case of Eya, I suggest that this indicates a difference of function between the Eya and D-Six4 proteins. Their null-mutant phenotypes may be similar in the muscle but their functions, albeit complementary, do not appear to be the same. Expression of Eya however results in the rescue of the ventral acute muscle 3. I suggest that this muscle requires either Eya or D-Six4 for its development but that in the absence of D-Six4, more Eya is necessary.

As discussed previously, no partial rescues were observed. However it seems unlikely that rescue of *D-Six4* function is either complete or none. Rather, the lack of scored partial rescued embryos is most probably the result of the imprecision in staging the embryos. Also using different muscle markers would help decipher whether there is partial rescue of *D-Six4* function. Additionally, (as also mentioned previously), the mutant phenotype itself is varied and therefore assessing partial rescue of *D-Six4* function is difficult. A more precise analysis of the *D-Six4*^{289} mutant phenotype will help resolve these difficulties.

5.3.2.3 What do the worsened embryos suggest?

Finally, in the muscle, I observe embryos displaying phenotypes of the ventral and lateral muscles that appear to be more severe than the *D-Six4*^{289} homozygous mutant phenotype when expressing under the control of a *twist*-GAL4 driver exogenous Optix. This phenotype is alleviated in the embryos misexpressing Optix under the control of a *twist*-GAL4 driver in a *D-Six4*^{289} heterozygous mutant, expressing one copy of wild-type D-Six4. I suggest that these results indicate an interference of Optix with D-Six4 cofactors and/or downstream targets. I suggest that this interference is protein dose-dependent. Future experiments would allow this hypothesis to be verified. Also, temporal control of expression of Optix would enable the putative identification of the cofactors with which it interacts and its downstream targets and would contribute to understanding not only how D-Six4 functions but also what are the common functional features between Optix and D-Six4. Future work could therefore include expressing Optix under the control of a heat shock promoter.

Importantly this phenotype requires precise characterisation using different muscle markers. The worsened phenotype was described as having additional missing muscles and an increase in unfused founder cells but a more precise description is required. As discussed above the *D-Six4*^{289} mutant phenotype is varied and precise characterisation of the mutant phenotype will help determine more specifically the increase in phenotypic severity observed when expressing Optix under the control of a *twist*-GAL4 driver.

5.3.2.4 What do the misexpression results suggest?

I showed that misexpressing *Drosophila Six* genes and *eya* under the control of a *twist*- GAL4 driver in the mesoderm of *D-Six4*^{289} heterozygous mutants resulted in different phenotypes depending on the protein and the tissue. Misexpression of either D-Six4, So or Optix has no phenotypic effect in the muscle. Together with the rescue experiments and assuming that the dose-dependent hypothesis is correct, I suggest that this indicates that these proteins when expressed in the same amounts as in the embryos which are homozygous mutant and display a mutant or worsened phenotype are outcompeted by endogenous D-Six4 in their putative competition for common cofactors. In the other case, when expressed in the same amounts as in the embryos which are homozygous mutant and display a wild-type phenotype, whether these proteins are outcompeted by endogenous D-Six4 or carry out its functions is unclear. Competition experiments would help answer this question. Overexpression of Eya results in phenotypes consistent with the previous ones observed in overexpression experiments of Eya.

5.3.2.5 What insights into the role of D-Six4 in the patterning of the mesoderm and the development of muscles do these results suggest?

In at least some aspects of somatic muscle patterning, the roles of *tinman* and *D-Six4* are complementary (Clark et al., 2006). Where *tinman* is required for the development of the visceral muscles and heart progenitors, *D-Six4* is required for the development of lateral and ventral muscles. *D-Six4* is important in the patterning of the ventro-lateral mesoderm. However, *D-Six4* also has more specific functions in the development of subsets of muscles as demonstrated by the consistent absence of VA3 for example in the *D-Six4*^{289} mutant. A more thorough characterisation of the *D-Six4*^{289} mutant phenotype and a precise analysis of which muscles are rescued when expressing So and Optix under the control of a *twist*-GAL4 driver will help decipher more about the role of *D-Six4* in the mesoderm and about conservation of function with the other two *Drosophila Six* genes.

5.3.3 What are the conserved/diverged functions of the Six proteins and Eya in the gonad?

5.3.3.1 What do the rescued embryos suggest?

I observed embryos displaying phenotypes of the germ cells indistinguishable from wild-type in *D-Six4*^{289} homozygous mutants when expressing exogenous D-Six4 under the control of a *twist*-GAL4 driver. Exogenous D-Six4 is the only protein showing the ability to fully rescue D-Six4 function in the gonad of *D-Six4*^{289} homozygous mutants. This indicates a potential for exogenous D-Six4 to carry out similar functions to endogenous D-Six4 involved in the patterning of the mesoderm and more specifically in the development of the gonad. This might include

downstream regulation of *Hmgcr* for example which is a direct transcriptional target gene of D-Six4 (Clark et al., 2006). Also, the wild-type number of germ cells populating the gonad is rescued when expressing D-Six4 in the *D-Six4*^{289} homozygous mutant embryos. Since Hmgcr expression is absent from SGPs in the *D-Six4*^{289} homozygous mutant, one could follow up all the results obtained (full and partial rescue) by assessing levels of *Hmgcr* mRNA in these embryos. This would be a direct molecular test of whether D-Six4 function has been restored or replaced.

5.3.3.2 What do the partially rescued embryos suggest?

I observed embryos displaying partially rescued phenotypes of the germ cells in *D-Six4*^{289} homozygous mutants when expressing exogenous D-Six4 and So respectively under the control of a *twist*-GAL4 driver. Both Six proteins show the ability to partially rescue D-Six4 function in the gonad of *D-Six4*^{289} homozygous mutants. The variation of phenotypes observed when expressing exogenous D-Six4 can be explained by the dosage-dependent hypothesis I explained previously. This indicates a potential for D-Six4 even when not expressed in the right amounts and So to carry out some of the functions involved in the patterning of the mesoderm and more specifically in the development of the gonad. The number of germ cells is lower than wild-type when attempting expressing exogenous So in a *D-Six4*^{289} mutant phenotype. This suggests that despite the ability for So to carry out some of the functions of D-Six4, there are some functions that are not rescued. This suggests a divergence between these two proteins in their ability to function in similar ways in the development of the gonad. In order to determine which are the functions that are

conserved, pull-down experiments with putative D-Six4 cofactors could be carried out in order to test the *in vitro* abilities of So to interact with these.

5.3.3.3 What do the mutant embryos suggest?

I observed embryos displaying phenotypes of the germ cells indistinguishable from mutants in *D-Six4*^{289} homozygous mutants when expressing under the control of a *twist*-GAL4 driver exogenous Optix and Eya, respectively. Both proteins are unable to rescue D-Six4 function in the gonad of *D-Six4*^{289} homozygous mutants. This indicates that Optix has functionally diverged from D-Six4 and is unable to carry out similar functions involved in the patterning of the mesoderm and more specifically in the development of the gonad.

Similarly to the phenotypes observed in the muscle, the mutant phenotypes observed when expressing exogenous Eya indicate its failure to rescue the *D-Six4*^{289} mutant phenotype and thereby carry out the same functions as D-Six4. I suggest that Eya and D-Six4 carry out complementary albeit different functions in the development of the gonad indicating either interactions with different cofactors or different interactions with the same cofactors.

5.3.3.4 What do the super mutant embryos suggest?

Finally, I observed embryos displaying a phenotype that I called super-mutant when expressing exogenous Optix under the control of a *twist*-GAL4 driver in *D-Six4*^{289} homozygous mutant embryos. These embryos showed an increase of germ cells coalescing into the gonad. I suggest that this increase is the result of additional germ cells coalescing. I suggest that Optix may contribute to inducing cell division of the

germ cells resulting in this increase. In wild-type embryos, not all germ cells coalesce into the gonad. Some are subject to cell death prior to their coalescence. This is thought to be a downstream effect of these germ cells not responding to the D-Six4 signal expressed in the SGPs (Clark et al., 2006). I suggest therefore that an alternative explanation for this increase of germ cells in the gonad is that Optix is able to rescue aspects of D-Six4 function resulting in the coalescence of germ cells and SGPs into two tightly formed gonads but additionally supersede D-Six4 function in promoting the signal that attracts germ cells that would normally not receive the D-Six4 signal into the gonad. These assumptions could be tested by staining with a caspase antibody to assess the amount of cell death or by staining for cell division markers or by looking at the levels of *Hmgcr* mRNA again possibly expecting higher levels than in wild-type.

5.3.3.5 What do the misexpression results suggest?

Misexpression of either D-Six4, Optix or Eya under the control of a *twist*- GAL4 driver in the mesoderm of *D-Six4*^{289} heterozygous mutants has no phenotypic effect in the gonad. Together with the rescue experiments, this indicates that neither Optix nor Eya are able to interact with D-Six4 cofactors and/or downstream targets in the development of the gonad. Expression of exogenous D-Six4 either interacts with the cofactors and downstream targets driving the development of the gonad or does not interfere with endogenous D-Six4 in its interaction with them.

I observed embryos displaying a mutant phenotype when misexpressing So under the control of a *twist*- GAL4 driver in the mesoderm of *D-Six4*^{289} heterozygous mutants. However, I also observed this phenotype in the control embryos which were

not theoretically expressing So but contained the [UAS-*so*] p-element. I concluded that this phenotype was the result a specific interaction between the presence of that p-element, possibly due to basal expression of So and the TM3 balancer. Balancers are carriers of many mutations. Indeed, mutations accumulate at the sites of inversions and also over time mainly because no recombination occurs to rid them of deleterious mutations. While determining the specificities of the interactions between the balancer and the [UAS-*so*] construct are of little interest, it appears critical to repeat these experiments with a different insertion line of So. Also, determining whether So is actually being expressed in these tissues may result in further understanding of these results.

5.3.3.6 What insights into the role of D-Six4 in the development of gonad do these results suggest?

Analysis of the hypomorph *D-Six4*^{131} mutant suggests that there are at least two mechanisms involved in the coalescence of germ cells, msSGPs and SGPs into the gonads (Clark et al., 2007). Some of the SGPs arising from the three initial clusters of SGPs associate even in the absence of *Hmgcr* together with one or two germ cells (Clark et al., 2007). This association is thought to be the result of short range signals not requiring *Hmgcr* and therefore not being D-Six4 dependent assuming that D-Six4 regulates *Hmgcr* (Clark et al., 2007). This process is known as compaction (Clark et al., 2007). This is different to the process of coalescence which requires long range signals from the SGPs to both the germ cells and the msSGPs to associate and coalesce into two tightly formed gonads. This process requires *Hmgcr* in the case of the germ cells but not in the case of the msSGPs which nonetheless respond to a long

distance signal as they do not associate in the compaction process (Clark et al., 2007). This is important when considering the variation observed in the mutant phenotype.

More quantitative and qualitative analysis of the *D-Six4*^{289} mutant gonad phenotype is required to assess whether the phenotypes classed as partial scatter in this thesis are perhaps the result of compaction as opposed to coalescence. This would contribute to understanding more precisely the role of D-Six4 in these two processes and whether these functions are conserved in the other Six proteins. This may include counting germ cells in the partial scatter and scattered phenotype and assessing the distances between them.

5.3.4 What are the conserved/diverged functions of the Six proteins and Eya overall?

I propose that the difference in rescue in the different tissues (muscle and gonad) is the result of the different conservation of the three Six proteins. I propose that depending on the levels at which they are expressed, D-Six4, So and Optix are able to rescue D-Six4 function in the muscle, and therefore have conserved the functions that D-Six4 carries out there, that is perhaps the ability to interact with common cofactors such as *tinman* and *Mef2* (chapter 1). I propose that So is able to rescue parts of D-Six4 function in the gonad, and therefore has conserved some the functions that D-Six4 carries out there, that is perhaps the ability to interact with *Hmgcr* or other cofactors of D-Six4. I propose that Optix is unable to rescue D-Six4 function in the gonad, and therefore has diverged from D-Six4 (chapter 1). Alternatively Optix not only rescues D-Six4 function but supersedes it.

These results are consistent with the fact that So and D-Six4 are more closely related than Optix. These results confirm the hypothesis that the Six proteins have functionally diverged to take on different functions specific to their subfamily.

Eya fails to rescue the *D-Six4*^{289} mutant phenotype both in the muscle and the gonad (except for VA3). This indicates that although Eya may be a cofactor of D-Six4 and carry out complementary functions it functions differently perhaps by binding differently to the same cofactors. A means of answering this question would be to carry out double Eya-D-Six4 mutants. If these assumptions are correct and Eya and D-Six4 do function together, I would expect the phenotype of these double mutants to be indistinguishable from the single null-mutants.

5.3.5 What are the limitations of this assay?

Functional conservation of these proteins was assessed by comparing the ability for these different proteins to rescue aspects of the *D-Six4*^{289} mutant phenotype. Conclusions made consequently are limited to the functions D-Six4 carries out in the development of the muscle and the gonad. Endogenous wild-type D-Six4 is also expressed in the developing head region and the central nervous system. Putative functional conservation of the Six proteins in these tissues may differ significantly from the results obtained in this study. Further work will include investigating this further.

A major limitation of this study is that only a single transformant line for each construct was tested. Consequently the percentages of rescue obtained may not be significant. Although the GAL4/UAS system has been widely used as a means of

expressing proteins exogenously and assessing their phenotypic effects, future work may include using a Six4-Six promoter fusion construct.

Chapter 6

Functional conservation and divergence between *Drosophila* Six proteins and their mammal orthologues

Functional conservation and divergence between Drosophila Six proteins and their mammal orthologues

6.1 Introduction

In the previous chapter, I showed that the three *Drosophila* Six proteins, D-Six4, So and Optix have diverged functionally and carry out functions specific to their subfamily. This is consistent with the results I obtained from protein sequence comparisons, whereby I showed that each family displayed features of putative functional conservation specific to that family. I hypothesised that these subfamily functions are conserved through evolutionary time and therefore consistent with previous suggestions, I propose that Six proteins within a subfamily are more closely related to each other than they are between Six proteins in a same species. I propose to test these hypotheses by assessing the functional conservation of *Drosophila* Six proteins and their mammal (mouse and human) orthologues. I propose to test the ability of mammal Six proteins to rescue aspects of the *D-Six4*^{289} mutant phenotype. I hypothesised that members of the Six4/Six5, Six1/Six2 and Six3/Six6 subfamilies respectively would show similar abilities to rescue the *D-Six4*^{289} mutant phenotype in the muscle and gonad as did D-Six4, So and Optix respectively.

For both the Six1/Six2 subfamily and the Six3/Six6 subfamily, one member of each has been more widely studied both in mouse and human, Six1 and Six3 respectively. While little is known about Six2 and Six6 in both human and mouse, mutations in *SIX1* are one of the causes of the branchio-oto-renal (BOR) syndrome in humans and *SIX3* has been found to be mutated in the homeodomain, in some patients with holoprosencephaly. In mouse, Six1 protein is involved in many steps of embryogenesis and Six3 plays a crucial role in early eye and forebrain development.

Additionally, the percentages of amino acid identity between the Six domains and the homeodomains of mSix1 versus mSix2 and mSix3 versus mSix6 range from 88% to 98%. I therefore focused my work on mouse Six1 and Six3. Six4 has been subject to extensive study in the mouse and is involved together with Six1 in the development of somites, the thymus, the kidney and some aspects of neurogenesis. Six5 is involved in the development and function of mesodermal tissues and brain in the mouse and is thought to be involved in myotonic dystrophy in humans. I consequently focused my study of the Six4/Six5 subfamily on mouse Six4 and mouse and human Six5. Throughout this chapter, to differentiate clearly between mouse and human *Six* genes and Six proteins, I precede each *Six* gene and Six protein with the letter m and h for mouse and human genes and proteins, respectively.

Transformant lines containing the p-elements [UAS-*mSix1*], [UAS-*mSix4*], [UAS-*mSix5*] and [UAS-*hSix5*] were not available before the start of my PhD. A transformant line containing the p-element [UAS-*mSix3*] was available for use (Zhu et al., 2002). My aims were to create transformant lines for all these constructs in order to then carry out similar experiments to the ones described in Chapter 5, enabling me to assess the conservation and divergence of function of these mammal proteins. While I did not achieve the goals I had set out to achieve, I report in this chapter on the work I carried out in designing and making the different constructs and in generating transformant lines. I experienced a number of critical experimental difficulties some of which could not have been anticipated at the start of this PhD. In this chapter, I also explain why I didn't use the p[UAS-*mSix3*] transformant line. The

work described in this chapter was an essential part of my training. The learning outcomes I developed have been an exciting part of my PhD.

6.2 Results

In order to generate transformant lines, I carried out work on four clones, *hSix5*, *mSix5*, *mSix4* and *mSix1*.

6.2.1 What were the results obtained whilst working with hSix5?

I obtained a clone of *hSix5* from Graham Hamilton in Glasgow (Ivan Clark personal communication). The clone came without a map or any information at all regarding it (Ivan clark, personal communication). I carried out some basic restriction digests to establish whether the insert was *hSix5* DNA and whether it was cDNA (no introns), gDNA, just the open reading frame or whether the untranslated regions were also present. From the sequence information available in Ensembl, I expected an insert of 2220 base pairs (bp) if it were cDNA and only the open reading frame and 3145bp if the untranslated regions were present and 4269bp if it were gDNA. From the restriction digests I carried out, I obtained a band of 2200bp and concluded that the insert was of the cDNA open reading frame only (figure 6.2.1.1).

I proceeded to sequence the insert in order to verify that it was the same sequence as the one in Ensembl. Not knowing what the vector was, I attempted to sequence the insert using combinations of primers that bind commonly used vectors. These include, T7, Sp6, M13. I also designed primers to anneal to the beginning and end of the insert based on the assumption that the insert sequence was identical to the one available in Ensembl.

Working with the *hSix5* DNA proved very difficult and despite efforts on Dr. Hamilton's behalf to retrieve information from his past co-workers about the construct, little was known about it and he reported that any work his lab had carried out with it had proven unsuccessful. Consistent with this, my numerous attempts to sequence the insert either resulted in poor quality traces or in poor quality sequences ridden with undetermined nucleotides. I also attempted to carry out PCR but despite optimising the conditions with different primers, different buffers and different polymerases and changing the amount of template and the extending temperature, I failed to amplify any product at all (table 6.2.1.1). I concluded that my failing to obtain a product from my PCR reactions was possibly due to the poor quality of the DNA. Ultimately, failing to obtain any significant results I resolved myself to being unsuccessful in generating a p[UAS-*hSix5*] construct.

The Six domain and homeodomain of mSix5 are respectively 99% and 100% identical in their amino acid sequence to hSix5. Eventually, I decided to focus my work on *mSix5* and *mSix4* instead.

ATGGCTACCTTGCCTGCGGAGCCGAGCGCG
GGGCCGGCGGCTGGGGGGGAGGCGGTGGCGGCGGCGGCGACCGAAGAGGAGGAGGAG
GAAGCGCGCCAGCTCTTGCAGACTTTGCAGGCGGCCGAGGGTGAGGCGGCGGCGGCGGCC
GGGGCCGGGGCGGGCGCAGCGGCTGCGGGAGCTGAGGGCCCGGGATCCCCGGGCGTCCCC
GGGTCGCCCCCGAGGCCGCTTCCGAACCGCCCACGGGCCTCCGCTTCTCGCCCGAGCAG
GTGGCGTGCGTCTGCGAGGCGCTGCTCCAGGCGGGCCACGCCGGCCGCTTGAGCCGCTTC
CTGGGCGCACTGCCCCCGGCCGAGCGCCTACGTGGCAGCGACCCGGTGTTGCGCGCGCGG
GCCCTGGTGGCCTTCCAGCGGGGCGAGTACGCCGAGCTCTACCGGCTACTCGAGAGCCGC
CCCTTCCCCGCCGCCCACCACGCCTTCCTGCAGGACCTCTACCTGCGCGCGCGCTACCAT
GAGGCCGAGCGGGCCCGCGGCCGCGCGCTTGGCGCAGTGGACAAGTATCGACTGCGCAAG
AAGTTCCCGCTGCCCAAGACCATCTGGGACGGCGAGGAGACAGTCTACTGCTTCAAGGAG
CGCTCCCGCGCAGCGCTCAAGGCCTGCTACCGCGGCAACCGCTACCCCACGCCGGACGAG
AAGCGCCGCCTGGCCACACTCACCGGCCTGTCGCTCACGCAGGTCAGCAACTGGTTCAAG
AACCGGCGACAGCGCGACCGGACCGGGGCCGGAGGCGGCGCGCCCTGCAAGAGCGAGTCT
GATGGGAATCCCACGACTGAGGACGAGTCCAGCCGAAGTCCTGAGGACCTGGAGAGAGGG
GCGGCCCCAGTGTCCGCCGAGGCCGCTGCCCAGGGCTCCATATTCCTGGCAGGGACCGGC
CCTCCCGCGCCTTGCCCGGCTTCCTCCTCCATCCTGGTGAACGGGAGCTTCCTGGCAGCC
AGCGGCTCCCCAGCAGTGCTCCTCAACGGGGGCCCCGTCATCATCAACGGCCTGGCCCTG
GGCGAGGCCTCCAGCCTGGGCCCGCTGCTGCTCACTGGGGGCGGGGGTGCCCCTCCACCG
CAGCCCAGCCCTCAGGGGGCCAGCGAGACCAAGACCTCTCTGGTCCTGGACCCTCAGACA
GGGGAGGTGCGGCTGGAGGAGGCTCAGTCGGAGGCCCCTGAGACCAAAGGGGCCCAGGTG
GCTGCTCCGGGACCAGCCCTTGGAGAGGAGGTCCTGGGGCCCCTGGCCCAAGTGGTGCCT
GGCCCCCCGACGGCTGCCACCTTTCCTCTGCCCCCGGGGCCAGTGCCTGCTGTGGCTGCC
CCACAAGTGGTACCGCTCTCCCCACCCCCGGGGTATCCCACGGGCCTGAGCCCCACCTCC
CCACTATTGAACCTGCCCCAGGTAGTACCCACCTCACAGGTGGTGACCCTGCCCCAGGCT
GTGGGGCCCCTGCAGCTGTTGGCAGCCGGGCCAGGCAGCCCTGTGAAGGTGGCAGCTGCA
GCAGGCCCTGCCAATGTGCACCTCATCAACTCCGGGGTGGGCGTGACTGCCCTGCAGCTG
CCTTCGGCCACTGCCCCAGGAAACTTCCTCCTGGCCAACCCTGTGTCTGGCAGCCCCATC
GTGACGGGTGTGGCCCTGCAGCAGGGCAAGATCATCCTCACCGCCACCTTCCCCACCAGC
ATGCTCGTCTCCCAGGTCCTGCCGCCAGCCCCCGGCCTGGCCCTGCCACTGAAGCCAGAG
ACGGCCATCTCCGTGCCTGAGGGAGGCCTCCCGGTGGCCCCCAGCCCTGCTCTCCCAGAG
GCTCACGCCCTAGGCACCCTTTCTGCACAGCAGCCACCCCCCGCCGCTGCCACCACCTCC
AGCACCAGCCTGCCCTTCTCCCCTGACTCCCCTGGCCTCCTGCCCAACTTCCCGGCGCCC
CCACCAGAGGGGCTGATGTTGTCACCCGCGGCCGTGCCTGTCTGGTCAGCAGGGCTGGAA
CTAAGCGCAGGAACAGAGGGGCTGCTGGAAGCGGAAAAGGGGCTGGGGACACAGGCCCCC
CACACCGTGCTGAGGCTGCCAGACCCCGACCCTGAGGGGCTGCTCCTGGGGGCCACCGCA
GGGGGTGAGGTTGACGAGGGGTTGGAAGCTGAGGCCAAGGTTCTGACCCAGCTCCAGTCG
GTGCCTGTGGAGGAGCCCTTGGAACTGTGA

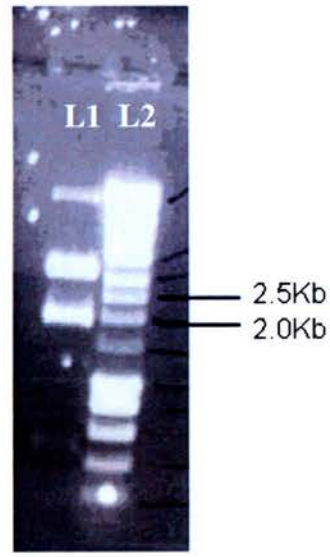


Figure 6.2.1.1: Coding sequence of *hSix5* (left) and a picture of the electrophoresis gel in which the digested *hSix5* was run (right). *hSix5* coding sequence is 2220bp long. *hSix5* was digested with EcoR1 and Xho1 and resulted in a product of about 2.2Kb (L1: digested *hSix5*, L2: marker). This confirms that the insert is the correct size and is likely to be the coding sequence.

Primers	Buffers	Polymerases	Template amount	Extending temperatures
*Annealing to the beginning and end of presumed sequence * Nested primers within the sequence *T7 and T3	Accompanying buffers of each Taqs, 12 buffers with differing pH, Mg and Salt concentrations.	Invitrogen Taq, Stratagene Taq long	5ng, 25ng,270ng	72C, 65C.

Table 6.2.1.1: The different PCR optimisations carried out in order to amplify *hSix5*.

6.2.2 What were the learning outcomes of working with hSix5?

I spent a considerable amount of time of my PhD trying to succeed in obtaining results with *hSix5*. I was still an inexperienced scientist at this point and was unsure of how the science community interacted. I now know that there are many other labs working on the same area of interest and that giving up with one source does not mean giving up altogether. There are other ways of obtaining *Six5* DNA. I obtained *mSix4* and *mSix5* DNA from other labs. Alternatively, I could have tried to amplify *hSix5* DNA from a sample of human gDNA.

6.2.3 What were the results obtained using mSix5 and mSix4?

I obtained *mSix5* and *mSix4* clones from the respective labs they had been used in for band-shift assays and in Western blot experiments (Sato et al., 2002; Spitz et al., 1998). These clones were accompanied by respective maps of the vectors they were in and the size of the inserts (figure 6.2.3.1 a and b). As for *hSix5*, I carried out basic restriction digests of the clones of both *mSix5* and *mSix4* to verify the information I had been sent. Based on this information, I obtained products of the expected size (1773bp and 2400bp respectively) and consequently designed primers in order to amplify through PCR the inserts and further subclone them (figure 6.2.3.2 a and b).

I designed my primers so that not only would they bind to the insert but also would add, upstream of the insert, a restriction enzyme site and a *Drosophila* ribosomal binding site (RBS). This is necessary because there are reported differences between *Drosophila* and mammalian RBS (Cavener, 1987). The *Drosophila* RBS consensus is CAAAAC whereas the vertebrate one is CANC (Cavener, 1987). Another consensus sequence that was successfully used in the lab is

GCCACC (Eve Hartswood, personal communication). Dr. Hartswood used this RBS to express plant genes in *Drosophila* and was successful in doing so. I therefore used this consensus in the design of my primers.

I carried out PCR experiments in an aim to amplify both *mSix5* and *mSix4*. I failed to obtain a product when amplifying *mSix5* but succeeded when amplifying *mSix4* (figure 6.2.3.3). In view of the difficulties I had experienced whilst working with *hSix5* DNA, I prioritised my objectives and in order to generate results focused my work on *mSix4* which at this early stage appeared promising. Because mSix4 and mSix5 only share 65% and 80% amino acid identity in their Six domains and homeodomains respectively, I intended to further pursue my work using *mSix5* but this was not achieved during the course of my PhD.

Cloning of the *mSix4* PCR product proved very difficult. I carried out numerous cloning experiments optimising for different conditions. I tried different amounts of PCR product in the ligation experiments (ranging from 10ng to 100ng). I tried ligating at different temperatures (room temperature, 4C) and for different periods of time (over night, 1hour, 2hours, 4hours). I optimised the method of transforming cells (heatshock, electroporation) as well as the method of making competent cells ($CaCl_2$, electrocompetent). I even tried ligating into pGEMT instead of pGEMTeasy (promega vectors commonly used for cloning PCR products, pGEMTeasy has an additional *EcoRI* restriction site on either side of the insert insertion). Finally, I succeeded in ligating the PCR product into pGEMT by electroporation. For this I used 70ng of PCR product and ligated at 4C overnight. I obtained three clones with inserts which I sequenced (figure 6.2.3.4 L1, L3 and L5). I obtained sequences with multiple errors that would cause non-synonymous codon

changes (3 in the first insert, 7 in the second and 4 in the third). To find the source of the errors, I sequenced the insert present in the starting vector PCR3 and found that it was identical to the one in Ensembl. I hypothesised that the PCR experiment had caused the incorporation of these mistakes and that the polymerase was the cause of this. I had indeed used a polymerase with no proof reading ability (NEB Taq polymerase). I tested this by carrying out PCR again using different polymerases with proof reading abilities (Taq long with a p*fu* proof-reading activity and Expand with Tgo proof reading activity). I successfully obtained a product which I cloned and sequenced. Unfortunately, all the inserts I sequenced were again mistake ridden and these were different from the previous mistakes. Time constraints did not allow further experiments to be completed in the time frame of this study.

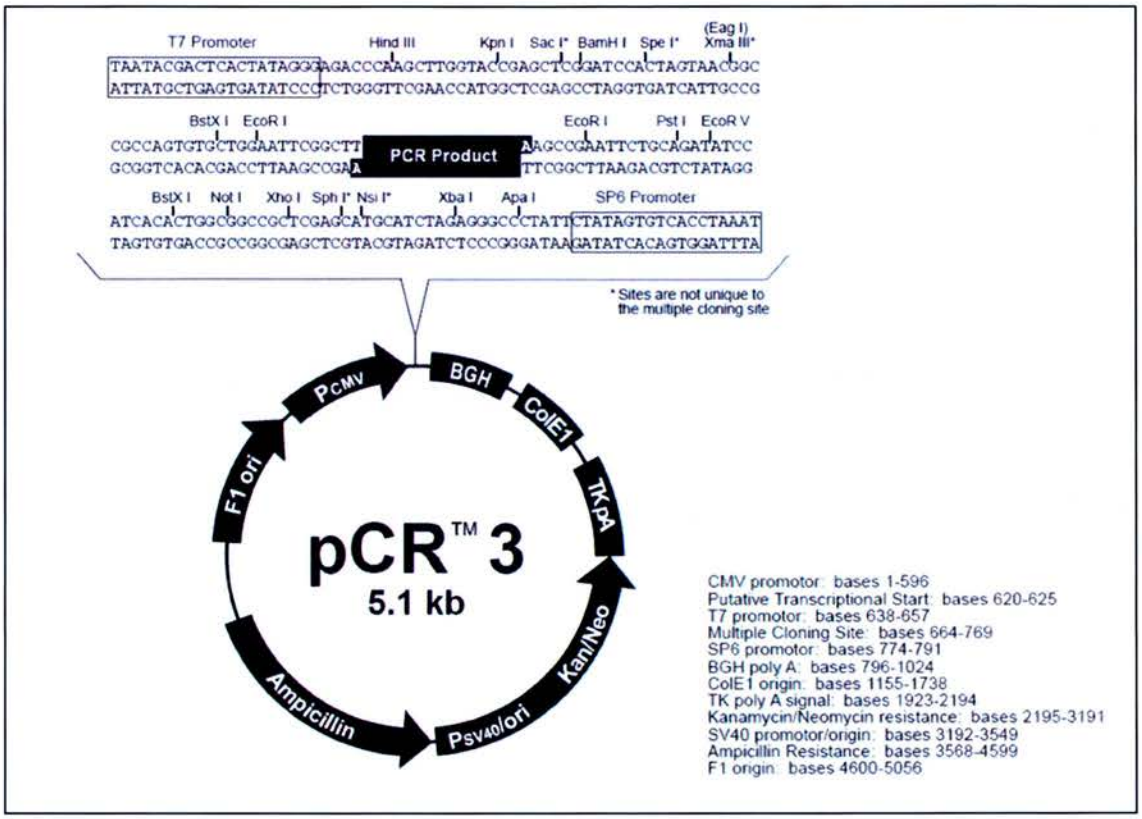


Figure 6.2.3.1a: A map of the pCR3 vector that the *mSix4* gene was initially cloned into. *mSix4* (2400bp) is cloned using the Bam H1-Not1 sites (Spitz et al., 1998).

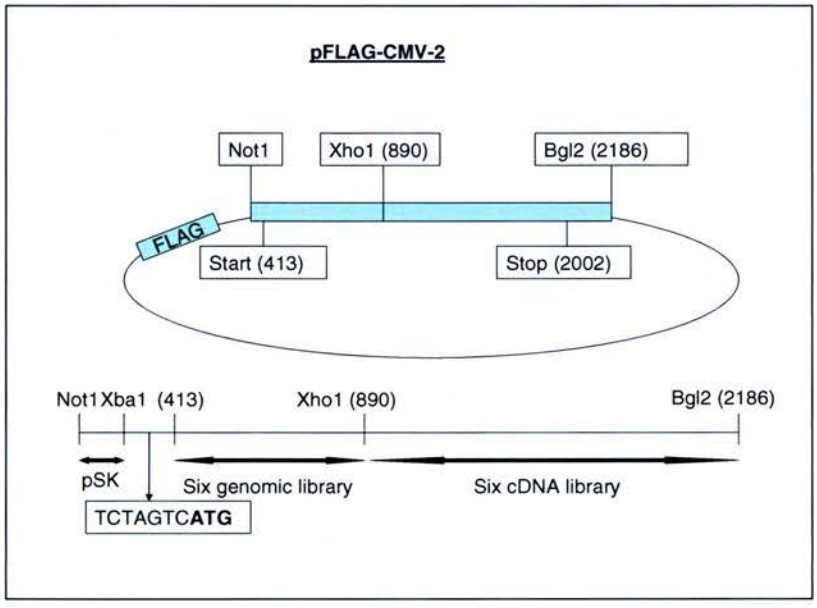


Figure 6.2.3.1b: A map of the pFLAG-CMV-2 vector that the *mSix5* gene was initially cloned into.

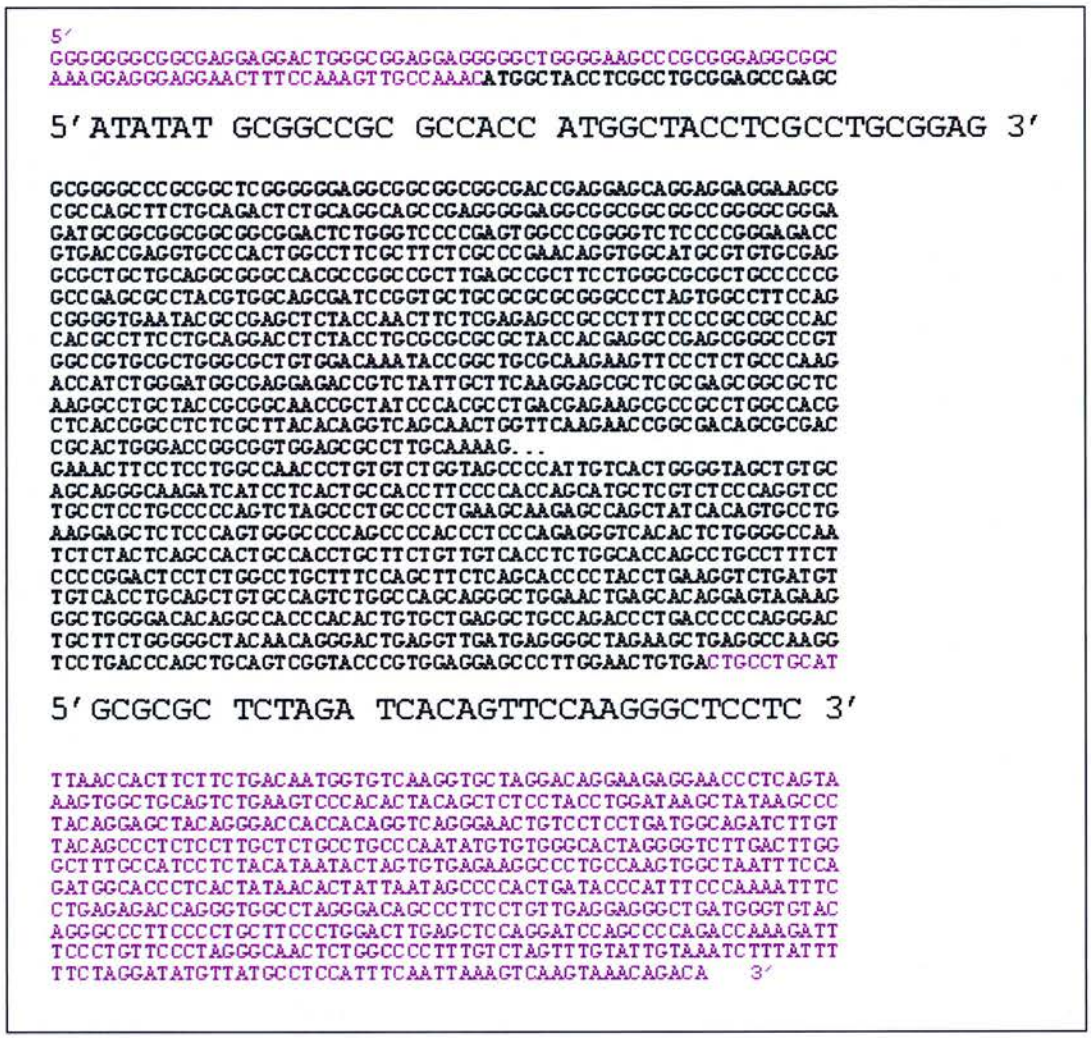


Figure 6.2.3.2a: DNA sequence of *mSix5*. The untranslated regions upstream and downstream of the coding sequence are shown in purple and the coding sequence in black. The three dots indicate the the end of the first exon and the start of the second exon. The primers shown in bigger font and flanked by 5' and 3' bind to the start and finish of the coding sequence to amplify *mSix5*. Each primer displays six "junk" nucleotides (added to approximate the AT/CG ratio to 50%) followed by a restriction enzyme site. The forward primer then displays the RBS followed by the sequence to which the primer anneals and the reverse primer just shows the sequence to which it anneals.

Figure 6.2.3.2 b: DNA sequence of *mSix4* cDNA. The primers flanked by 5' and 3' bind to the start and finish of the coding sequence to amplify *mSix4*. Each primer displays six "junk" nucleotides (added to approximate the AT/CG ratio to 50%) followed by a restriction enzyme site. The forward primer then displays the RBS followed by the sequence to which the primer anneals and the reverse primer just shows the sequence to which it anneals.

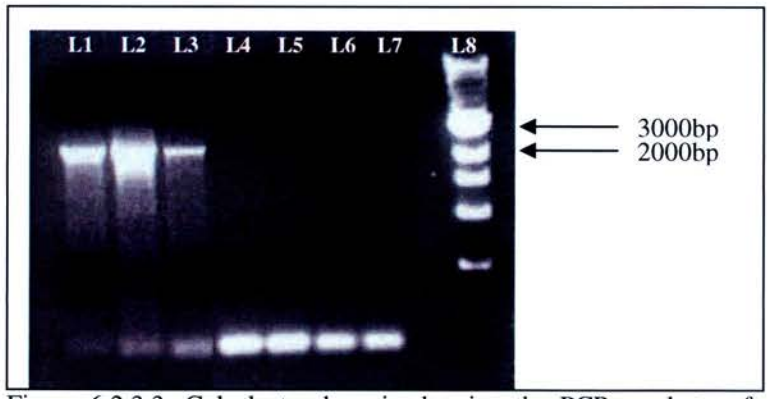


Figure 6.2.3.3: Gel electrophoresis showing the PCR products of *mSix4* and *mSix5* reactions. L1, L2 and L3 show a *mSix4* product of approximately 2400bp. L4, L5, L6 and L7 are void of any *mSix5* product. L8 is the marker.

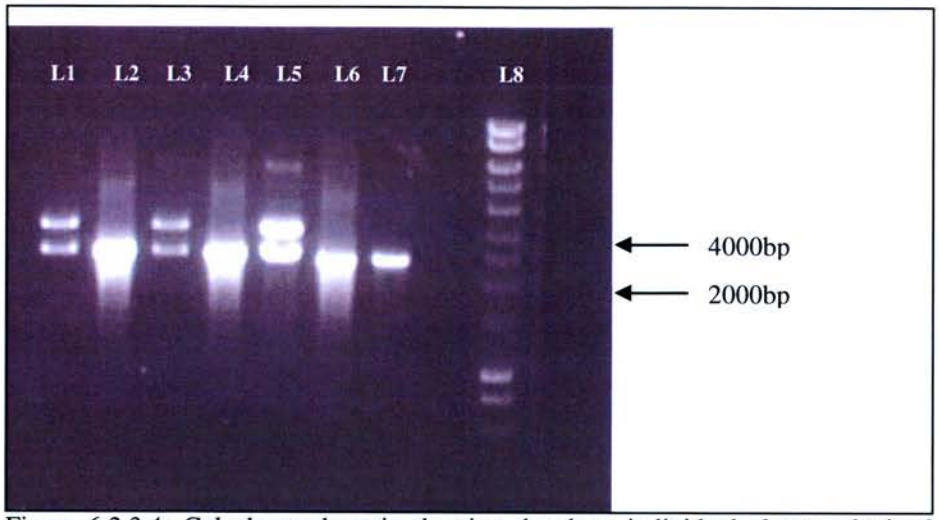


Figure 6.2.3.4: Gel electrophoresis showing the three individual clones, obtained from the cloning of the *mSix4* PCR product, digested with the restriction enzymes integrated in the primers alongside the *mSix4* PCR products. L1, L3 and L5 show two products: the bigger one is the vector mSix4 was successfully ligated into, the smaller is the *mSix4* insert. L2, L4, L6 and L7 show the *mSix4* PCR product. L8 is the marker.

6.2.4 What were the learning outcomes using mSix4?

Designing a molecular construct is theoretically very easy. You start by obtaining a source of DNA insert which you perform a restriction digest on and sequence in order to be certain of what you have. You design primers to amplify the DNA by PCR. These primers have add-on features such as the *Drosophila* ribosomal binding site I discussed. You carry out a PCR on the DNA you wish to amplify and clone the PCR product into a first vector. You then sequence the insert again in order to ensure that the polymerase did not generate any mistakes during the PCR and then you clone your insert into a second vector which is appropriate for generating a p[UAS-*gene*] construct. This construct is now ready to be injected into *Drosophila* embryos.

In practice, it is all very different. In designing the *mSix4* construct, I was faced with hurdles every step along the way. While I attempted to trouble shoot each difficulty I was faced with, ultimately due to time constraints I was unable to succeed in generating a *mSix4* construct. Eventually my greatest success was the realisation of practical time constraints that need to be considered when designing experiments. Also my understanding of molecular genetics and its applications was considerably increased by all these tribulations.

6.2.5 What were the results obtained using mSix1?

I obtained the *mSix1* clone from the same lab as the one where I obtained *mSix4* (Sato et al., 2002; Spitz et al., 1998). Both *mSix4* and *mSix1* were cloned into the same vector (pCR3) so the accompanying map was the same as previously shown (figure 6.2.3.1a). *mSix1* (890bp) is cloned in the pCR3 vector using the Bam H1-Not1 sites (Spitz et al., 1998). As previously described, I carried out basic restriction

digests of the clone to verify the information I had been sent. Based on this information, I obtained a product of the expected size (890bp) and consequently designed primers as previously described in order to amplify through PCR the insert and further subclone it (figure 6.2.5.1). The insert was successfully amplified (figure 6.2.5.2), gel purified and cloned into pGEMTeasy. This was successful at the first attempt (figure 6.2.5.3). I subsequently sequenced the insert to ensure that the PCR had not inserted sequence random errors and to ensure that the RBS I had incorporated into the primers was present. The sequence obtained was identical to the cDNA sequence deposited in the Ensembl database.

Prior to the design of the primers, I carried out a restriction analysis of the vectors and the insert involved in the experiments I was to carry out. I selected restriction enzymes that were absent in the insert and present in the multiple cloning sites of the vectors and added them to the primers. I then isolated the insert from the vector using the restriction sites, I had added into my primers, by gel purification and ligated it into the pUAST vector. Again, this was an immediate success. Although theoretically unnecessary as no mutation should have occurred during the ligation procedure, I sequenced the insert as a precautionary step as errors may have occurred during the UV exposure during the gel purification step. Upon the result of it being identical to the Ensembl sequence, I injected the plasmid DNA into *Drosophila* embryos.

Figure 6.2.5.1: This shows the untranslated regions upstream and downstream of the coding sequence and the coding sequence of *mSix1*. The primers shown in bigger font and flanked by 5' and 3' bind to the start and finish of the coding sequence to amplify *mSix1*. Each primer displays six "junk" nucleotides (added to approximate the AT/CG ratio to 50%) followed by a restriction enzyme site. The forward primer then displays the RBS followed by the sequence to which the primer anneals (bold) and the reverse primer just shows the sequence to which it anneals (bold).

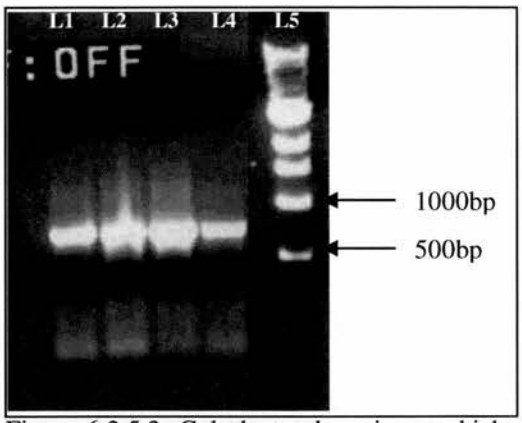


Figure 6.2.5.2: Gel electrophoresis on which the *mSix1* PCR products were run. L1, L2, L3 and L4 show a product of approximately 890bp indicating that *mSix1* was successfully amplified by PCR. L5 is the marker.

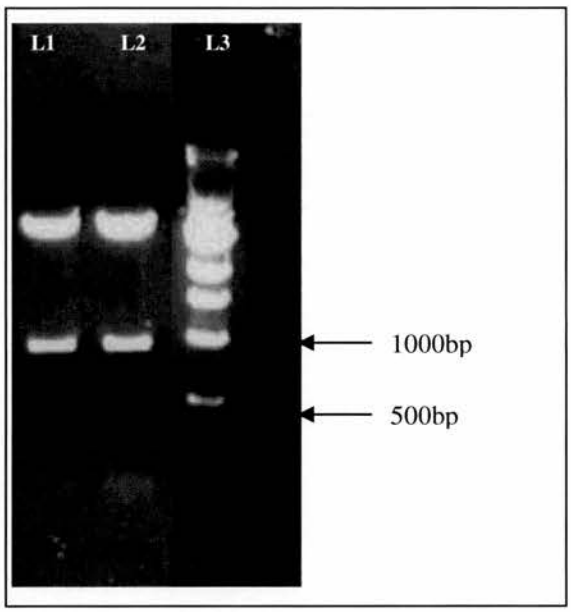


Figure 6.2.5.3: Gel electrophoresis showing the two individual clones, obtained from the cloning of the *mSix1* PCR product, digested with the restriction enzymes integrated in the primers. L1 and L2 show two products: the bigger one is the pGEMT vector *mSix1* was successfully ligated into, the smaller is the *mSix1* insert. L3 is the marker.

6.2.6 How did I generate transformant lines?

P[UAS-*mSix1*] is the only construct I successfully managed to generate. I injected it into 350 *Drosophila* Δ2-3 embryos as described in chapter 8. *Drosophila* Δ2-3 flies contain a transposase which enables the hopping of the inserted p-element. When the p-element is inserted in a region of the genome where it may be expressed, this can be observed phenotypically; white eyed flies display mosaic red/white eyes. I obtained 95 transformant lines (27% success rate), that is 95 flies which displayed mosaic eyes. These mosaic flies were repeatedly crossed with pure-white eyed flies until the progeny displayed pure red eyes indicating that the Δ2-3 transposase p-element had been lost and the inserted p-element of interest, in this case p[UAS-*mSix1*] was stably inserted in a region of the genome where it may be expressed. I then carried out a number of crosses to determine the chromosomal locations of the different insertions.

6.2.6.1 How did I map the insertion onto the X chromosome?

I first carried out a cross which enabled me to assess whether the p-element had inserted into the X chromosome (figure 6.2.6.1.1). For lines in which the p-element was indeed established to be on the X chromosome, I carried out a number of crosses to generate a stable stock (figure 6.2.6.1.2).

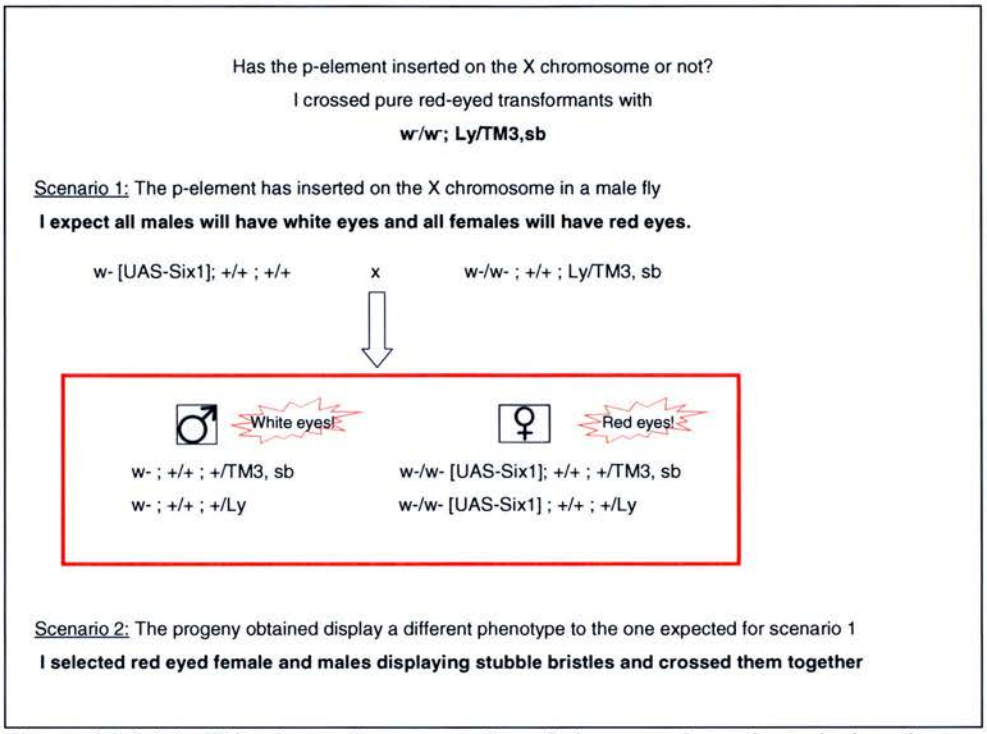


Figure 6.2.6.1.1: This shows the crosses I carried out to determine whether the p-element had inserted on the X chromosome or not.

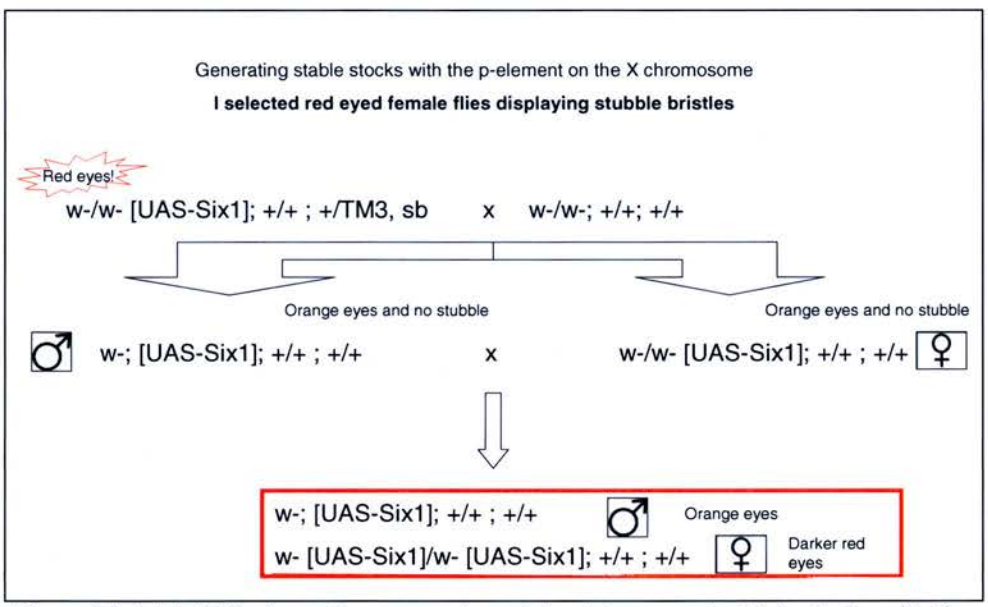


Figure 6.2.6.1.2: This shows the crosses I carried out to generate stable stocks with the p-element inserted on the X chromosome.

6.2.6.2 How did I map the insertion onto the second and third chromosomes?

For insertions that did not map to the X chromosome, I then had to assess whether the p-element had inserted onto the 2nd or 3rd chromosome. The crosses required for this are shown (Figure 6.2.6.2.1). Once I had established what chromosome the p-elements had inserted on, I carried out a number of crosses to generate stable stocks (Figure 6.2.6.2.2).

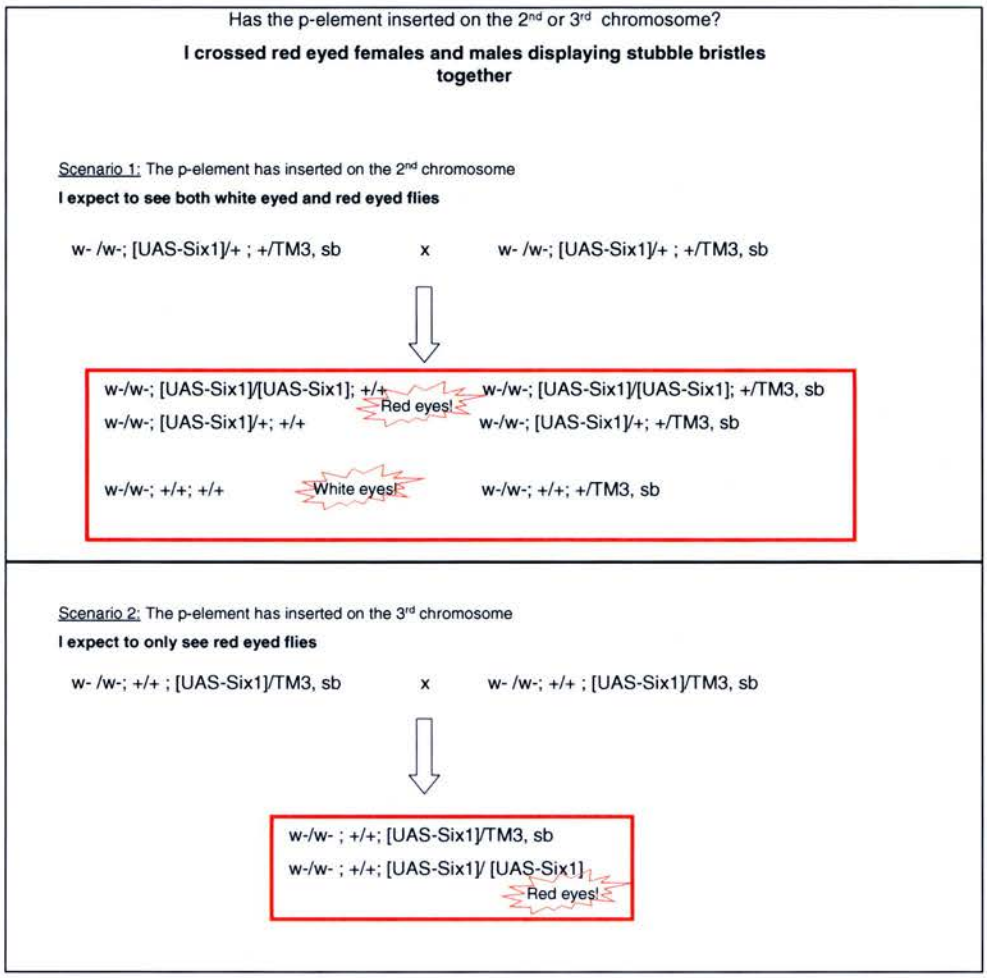


Figure 6.2.6.2.1: This shows the crosses I carried out to determine whether the p-element had inserted on the 2nd or 3rd chromosome.

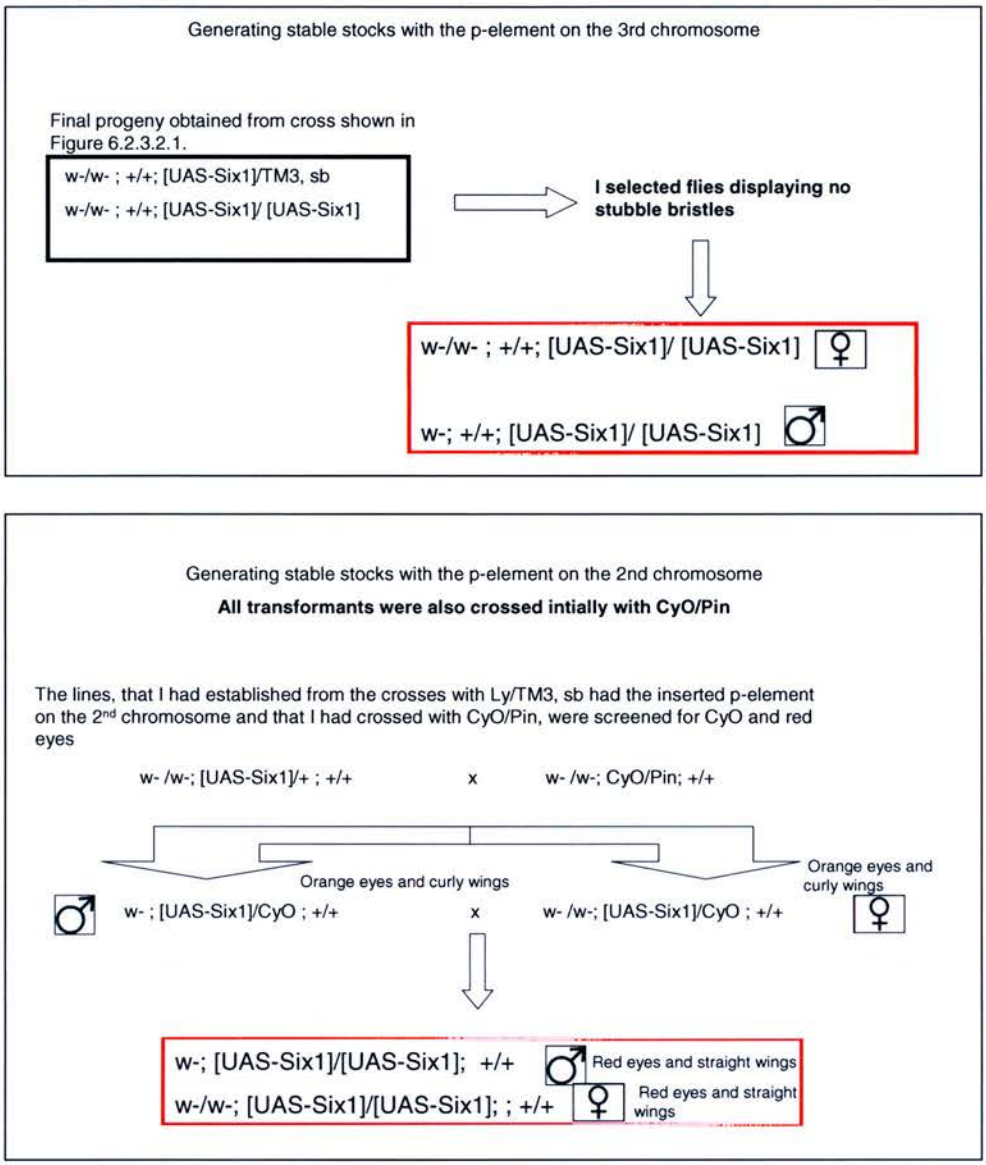


Figure 6.2.6.2.2: This shows the crosses I carried out to generate stable stocks with the p-element inserted on the 2nd and 3rd chromosomes.

6.2.7 Using the p[UAS-Six1] and p[UAS-Six3]

Once the insertions were mapped, 18 stocks were available for use (table 6.2.7.1). I used the six different transformant lines with homozygous p[UAS-*mSix1*] on the 2nd chromosome and further crossed them to change their genetic background to a *D-Six4*^{289} mutant one, balanced with a TM3 p[*Ubx-LacZ*] balancer (figure 6.2.7.1).

Whilst these crosses were in process, I also crossed each of the six transformant lines with homozygous p[UAS-*mSix1*] on the 2nd chromosome with flies containing the p-element p[GMR-GAL4]. GMR will drive the expression of GAL4 in the eye. Overexpression of So in the eye results in a major disruption of eye development which is phenotypically visible (Cheyette et al., 1994). The assumption was that because of the similarities between mSix1 and So, they are both part of the same subfamily and are 85% and 93% identical in their Six domain and homeodomain respectively. Expressing mSix1 under the control of a GMR-GAL4 driver may result in a similar phenotype. The expectation was that expressing mSix1 in the eye would have a disruptive effect. This was done in order to verify which of the transformant lines containing the p[UAS-*mSix1*] element were successfully expressing mSix1. None of the flies of the six transformant lines showed any visible phenotypic effect when expressing mSix1 in the eye under the control of a GMR-GAL4 driver. This suggests that mSix1 expression in the eye has no visible disruptive effect. Perhaps mSix1 and So are not so similar after all. Alternatively, the mSix1 is not successfully being expressed. Since all six transformant lines showed the same negative result, it was hypothesised that there may be a problem with the construct. Also, although I was able to successfully generate flies with the genotype p[UAS-*mSix1*]/p[UAS-*mSix1*]/, *D-Six4*^{289} TM3 p[*Ubx-LacZ*] and although these flies

reach adulthood, they are very poorly and die within 24 hours. In view of these results and due to financial constraints, I was unable to carry out any further experiments with them. Had I crossed them with the flies carrying the p[*twist*-GAL4] driver I would either have observed a different phenotype to the mutant one or the same phenotype as the mutant one. In the latter case, I would have needed to further test using a mSix1 antibody whether the mSix1 protein was being expressed and if so would have concluded that it neither rescued D-Six4 function in the mesoderm nor disrupts the development of the eye when misexpressed in that tissue.

Additionally, I had flies with heterozygous p[UAS-*mSix3*] on the 2nd chromosome (homozygous p[UAS-*mSix3*] flies are lethal (Zhu et al., 2002). I carried out crosses to change their genetic background to a *D-Six4*^{289} mutant one, balanced with a TM3 p[*Ubx-LacZ*] balancer (figure 6.2.7.2). The experiments involving these flies had not been a priority throughout this PhD mainly because the Six proteins belonging to the Six3/Six6 subfamily are more distantly related to the other Six proteins. I therefore only started work on this towards the last few months of my PhD. Unfortunately, I was unable to generate flies with the genotype p[UAS-*mSix3*]/*CyO*, *D-Six4*^{289}/ TM3 p[*Ubx-LacZ*] in the time frame of this study and this work could not be continued.

w- [UAS-*mSix1*]/ w- [UAS-*mSix1*] ; +/+; +/+	w-/w-; +/+; [UAS-*mSix1*]/TM3, *sb* w-/w-; +/+; [UAS-*mSix1*]/ [UAS-*mSix1*]	w-/w-; [UAS-*mSix1*]/ [UAS-*mSix1*]; *D-Six4*^{289}/ TM3 p[*Ubx-LacZ*] w-/w-; [UAS-*mSix1*]/ *CyO*; *D-Six4*^{289}/ TM3 p[*Ubx-LacZ*]
6	6 (mixture of both genotypes)	6 (mixture of both genotypes)

Table 6.2.7.1: Stocks of flies generated with the UAS-*mSix1* p-element, their genotypes and the number of transformant lines available.

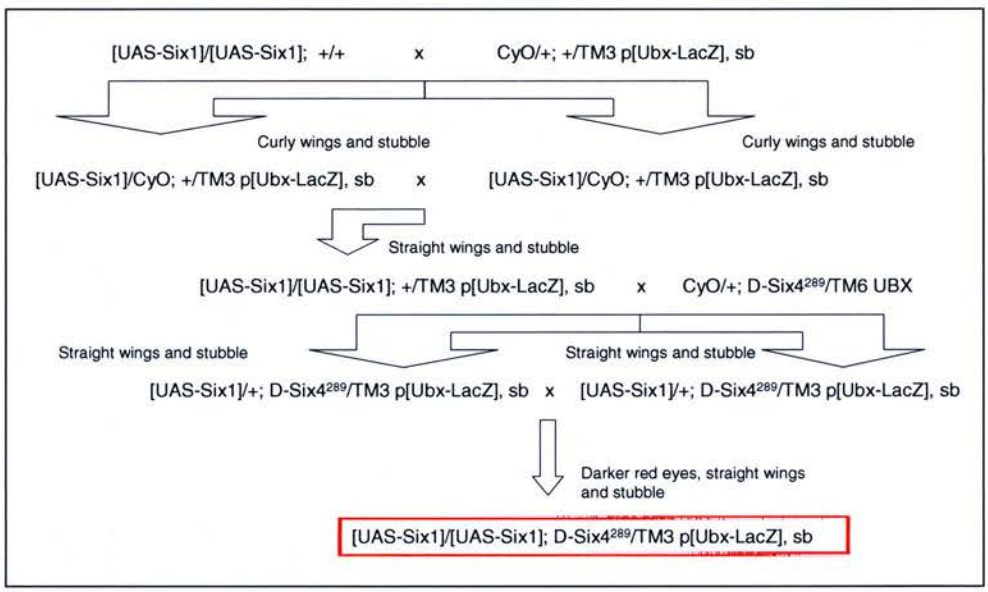


Figure 6.2.7.1: Crosses I carried out in order to obtain flies with the genotype shown in the red box. These were crossed with males of the same genotype and kept as a stock.

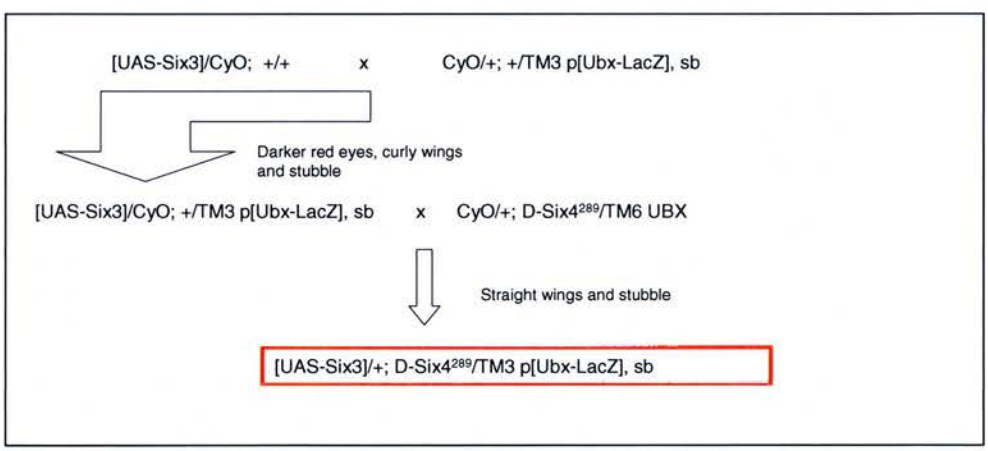


Figure 6.2.7.2: Crosses I carried out in order to obtain flies with the genotype shown in the red box. These were never obtained as a contamination occurred and through time constraints the project was stopped.

6.3 Discussion

In this chapter I reported on the molecular work I carried out in order to construct four different transformant lines, p[UAS-*hSix5*], p[UAS-*mSix5*], p[UAS-*mSix4*] and p[UAS-*mSix1*]. The work I carried out using *hSix5* DNA clone was unsuccessful and I believe that there is no point in pursuing future work with it since the cause is original and I believe the DNA clone I was sent is the problem. The work I carried out using *mSix5* DNA was also unsuccessful but mainly because it did not generate immediate results and because of time constraints I chose to focus my work on *mSix4* and *mSix1* DNA instead. I believe that more work with *mSix5* DNA may generate results and future work will involve optimising the PCR conditions so as to generate a product and continue from there. I successfully amplified *mSix4* DNA and cloned it. However the PCR experiment generated mistakes. More work will involve optimising the PCR conditions so as to minimise these mistakes. I successfully designed and generated the p[UAS-*mSix1*] construct and used it to make transformant lines. Time constraints and the fact that flies with the genotype p[UAS-*mSix1*]/p[UAS-*mSix1*]/, *D-Six4*^{289} TM3 p[*Ubx-LacZ*] die have resulted in these flies not being used. Future work will involve crossing six different transformant lines with flies containing the [*twist*-GAL4] construct in order to assess whether mSix1 is able to rescue D-Six4 function in the muscle and the gonad. Similarly, due to time constraints and experimental difficulties, I was unable to use the p[UAS-*mSix3*] transformant line. Future work will involve crossing it also with flies containing the [*twist*-GAL4] construct in order to assess whether mSix3 is able to rescue D-Six4 function in the muscle and the gonad.

Chapter 7

Discussion

Discussion

7.1 Summary of results

In this thesis, I analysed the functional conservation and divergence of the Six family of homeodomain transcription factors.

Firstly, I established criteria defining the Six domains and the homeodomains of Six proteins. I compared the protein sequence of the Six domains and homeodomains of Six proteins in the *Drosophila* genus, *Anopheles gambiae*, *Apis mellifera*, mouse and human and showed that the Six domains of these proteins are characterised by a specific decapeptide sequence and the homeodomains by a tetrapeptide. I also showed the conservation of previously known invariant amino acids in these species. I showed that the previous classification of the Six proteins into three subfamilies was supported by amino acid differences in the decapeptide relative to each subfamily and conserved in the members of these subfamilies. Additionally, I showed the conservation of the previous characterised homeodomain subfamily specific tetrapeptide. These results support and complement the previous characterisation of the Six proteins and their previous classification into three subfamilies.

I then established the phylogeny of the Six proteins and showed that these proteins were conserved from sponges to humans. This established the *Six* genes as an ancestral family of transcription factors, whose functions are conserved across evolutionary time.

In chapter 4, I combined the work I had carried out in the first two result chapters and showed that fourteen amino acids were unique respectively in each of the three *Drosophila* proteins and that these were conserved through evolutionary

time. I proposed that this conservation was functionally significant. In particular, I isolated four amino acids which I proposed as candidate amino acids for future analysis.

In Chapter 5, I discussed the results I obtained from rescue experiments whereby under the control of a *twist*-GAL4 driver, I expressed the three *Drosophila* proteins in the mesoderm of *D-Six4*^{289} mutant embryos in order to assess whether these proteins could rescue D-Six4 function. I showed that while exogenous D-Six4 can rescue D-Six4 function both in the muscle and in the gonad, So can only partially rescue D-Six4 function and Optix cannot in the gonad and only very inefficiently in the muscle. The varying capabilities of D-Six4, So and Optix rescue D-Six4 function indicate varying levels of functional conservation. This supports the previous hypotheses that the Six1/Six2 and Six4/Six5 subfamilies of which So and D-Six4 are members respectively are more closely related than they are to the Six3/Six6 subfamily of which Optix is a member. I also showed that expressing Eya under the control of a *twist*-GAL4 driver in the mesoderm of *D-Six4*^{289} mutant embryos does not rescue D-Six4 function supporting previous suggestions that although they may function together, D-Six4 and Eya carry out complementary but different functions. These experiments were complemented by misexpression experiments which indicated that misexpression of D-Six4 and Optix under the control of a *twist*-GAL4 driver in the mesoderm of *D-Six4*^{289} heterozygous mutant embryos has no effect. Consistent with previous results misexpression of Eya results in the duplication of the VA3 muscle. Misexpression of So however results in some cases in a mutant phenotype in the gonad. Together, these results support the three subfamily classification established previously and corroborated in Chapter 2.

Finally, in Chapter 6, I reported the work I carried out in generating constructs of [UAS-mouse/human Six proteins]. Mouse and human DNA proved more difficult than expected to work with and although I successfully generated a p[UAS-*mSix1*] construct, I was unable to carry out any experiments with it.

7.2 Did I achieve what I had set out to?

The Six family of transcription factors are involved in many aspects of development in many different species. Understanding how they function sheds light on these developmental processes. In order to gain further understanding of these Six proteins, the aim of this PhD was to assess whether different members of the Six family of transcriptional factors have distinct functions that have remained conserved over time.

The Six proteins have been classified into three subfamilies. I hypothesised that the different subfamilies of Six proteins display functional differences that are conserved within members of the same subfamily even between different species.

I showed with caveats that the three *Drosophila* Six proteins have functionally diverged. Optix is unable to carry out D-Six4 functions when expressed under the control of a *twist*-GAL4 driver in the mesoderm of $D\text{-}Six4^{289}$ mutant embryos and So can only partially compensate for loss of *D-Six4* in these conditions. This indicates a divergence of these proteins' functions consistent with my hypothesis.

I hypothesised that these functional differences are conserved through evolutionary time. Unfortunately during the course of this PhD, I was unable to test

out this hypothesis and future work leading on from my PhD will involve verifying this hypothesis.

7.3 What immediate future work can be carried out?

From the bioinformatics chapters the immediate future work would be to verify the functional significance of the features of conservation I identified by mutating the six amino acids I selected and assessing through rescue experiments similar to those carried out in this PhD, the effect on the function of these proteins.

A limitation of this PhD was that the results obtained from the rescue experiments were those observed in one transformant line. It may be that the results observed are specific to the insertion. This could be verified by carrying out the same experiments using other transformant lines. This would also allow for the verification and validation of the percentages of rescue obtained in this thesis.

Function of these proteins in mammals can be assessed by carrying out the crosses with the flies I made carrying the p[UAS-*mSix1*] and the p[UAS-*mSix3*] constructs. Additionally, I recently found out that a transformant line with the p[UAS-*mSix4*] construct is available and I have requested it (Justin Kumar, personal communication). Upon receipt of this construct and following a few crosses to change the genetic background to a *D-Six4*^{289} mutant one, whether these proteins are functionally conserved can be assessed. I would expect that mSix4 would rescue the D-Six4 function and that mSix3 would not. Indeed, mSix4 belongs to the same subfamily as D-Six4 and the functions they carry out in their respective species are comparable. D-Six4 is involved in the mesoderm and the development of the gonads and the muscle and mSix4 contributes in mouse to the development of somites.

mSix3 belongs to the same subfamily as Optix which I showed was unable to rescue the D-Six4 function when expressed in the mesoderm under the control of a *twist*-GAL4 driver. Also mSix3 is involved primarily in the development of the brain and the visual system in mouse. I therefore think it unlikely to be able to carry out the same functions as D-Six4.

As far as mSix1 is concerned, my expectations are very mixed. Theory would lead me to suggest that mSix1 would be unable to rescue D-Six4 function. Indeed, mSix1 is part of the same subfamily as So and thus I would not expect any more rescue than that observed when expressing So in the mesoderm under the control of a *twist*-GAL4 driver. However, experimental results showing the extensive role of mSix1 together with mSix4 in the development of many organs in mouse lead me to suggest that id mSix4 is able to rescue D-Six4 function then mSix1 may very well too. These results are the most expected as they would contribute significantly in determining whether the three subfamily classification is appropriate when discussing function of the Six protein family of homeodomain transcription factors.

7.4 Future work

As discussed in previous chapters, functional conservation studies in eye development between So and Optix have shown that while the N-terminal regions of So and Optix appear to be completely dispensable and interchangeable, replacement of the So C-terminal tail with the C-terminal tail of Optix appears to have an inhibitory effect (Weasner, Salzer et al. 2007). Also, deletion or replacement (with Optix SD) of the So SD regions rendered the modified So proteins incapable of restoring eye development (Weasner, Salzer et al. 2007). A protein lacking the HD of

So failed to rescue *so* mutant however expression of a protein in which the HD domain of So was replaced with the HD of Optix partially rescued the same *so* mutant (Weasner, Salzer et al. 2007). This reflects a partial conservation in the DNA binding specificity of the So and Optix homeodomains (Weasner, Salzer et al. 2007). An expressed a protein in which both the SD and HD domains of So were replaced with the corresponding domains of Optix did not display any restoration of eye development (Weasner, Salzer et al. 2007).

I believe it would be both interesting and informative to carry out similar experiments in the mesoderm and assess whether the functional conservation observed between Optix and So in the development of the eye is maintained in the mesoderm. This would contribute to understanding what the functional specificities between these proteins are namely in terms of assessing whether their conservation is general or is specific to certain co-factors and binding sites. These domain swap experiments could also involve swapping the domains of So and Optix proteins with the domains of D-Six4. Rescue experiments would then help determine the specificity of each domain and would give a clearer understanding of their respective importance in the overall functioning of the protein.

Recently, cofactors of So and Optix have been identified in the eye, Sine oculis binding partner (Sbp) and Optix binding partner (Obp) respectively (Kenyon, Yang-Zhou et al. 2005). These transcription factors and their differential abilities to interact with either So or Optix were shown through rescue experiments to be an important mechanism in determining significant functional differences between these two proteins during eye development (Kenyon, Yang-Zhou et al. 2005). I have tried to assess whether these proteins are expressed in the mesoderm as their presence

there would indicate that their proven interactions with So and Optix might not be limited to just those two Six proteins but that instead they may also interact with D-Six4. Very little is known about Sbp and Obp and mutants are not yet available. I believe that more work on these proteins will shed light on the molecular functions of the Six proteins.

In this thesis, I have discussed the functional conservation and divergence of the Six proteins and have assessed it by assessing rescue of D-Six4 protein in the muscle and the gonad by histochemically staining embryos. Interesting future work might include assessing whether the rescued mutant embryos show an ability to hatch. Future work into the physiology of these mutants may generate interesting results. I have assessed functional rescue phenotypically. It may be that actual function is limited. In collaboration with Phillippa Newby, I carried out some preliminary physiological analysis of the D-Six4 hypomorph mutant that suggested that the ability of mutant embryos to hatch was reduced in comparison to wild-type. Similar experiments could be carried out using the *D-Six4*^{289} mutants. Particularly, it would be interesting to compare the ability to hatch of embryos rescue whilst expressing exogenous D-Six4 in comparison to those partially rescued whilst expressing exogenous So. These experiments would further information on the functional specificities of these proteins.

Finally, a critical missing asset in the study of the Six family is a crystal structure of the protein domains. Generating a crystal structure of the Six domain and the homeodomain of these proteins would enable us to understand what contacts the protein makes with other proteins and DNA binding sites thus highlighting putative functionally important amino acids. This will inform us on the importance of

function of the signature sequences that I identified in Chapter4, for instance the Six domain decapepetide and the homeodomain tetrapepetides. In collaboration with Veronique Shulten, I carried out some preliminary work towards the expression of the Six domain and the homeodomain. Future work furthering this would help decipher the crystal structure of the Six proteins' Six domains and homeodomains.

The Six family of transcription factors are a critical family of proteins and more work carried out on their study will help us understand the molecular basis of development.

Chapter 8

Materials and Methods

Materials and Methods

8.1 Materials

8.1.1 Media

8.1.1.1 Bacterial Media

Luria Broth (LB)

Bacto tryptone (BD), 10g; Bacto yeast extract (BD), 5g; NaCl, 5g; per litre adjusted to pH 7.2

Luria Agar (L-agar)

Luria broth with 15g/l Bacto agar (BD). Ampicillin (Penbritin, Beecham Research) was added to LB and L-agar to a final concentration of 100μg/ml where indicated.

SOC Buffer

LB with 3.6g/l glucose, 0.1M $MgSO_4$ and 0.1M $MgCl_2$

2xTY Broth

Bacto tryptone (Difco), 16g; Bacto yeast extract (Difco), 10g; NaCl, 10g; per litre adjusted to pH 7.4

8.1.1.2 *Drosophila* Media

Dundee Fly Food

443g brewers yeast, 714g maize, 57g live yeast, 786g glucose, 27g nipegin, 107g agar 32ml propionic acid up to 10L with water.

Grape juice agar

Bacto agar (Difco), 205g per 100ml pure grape juice

8.1.2 Chemicals

Chemicals were supplied by Fisher, New England Biolabs, Promega, Roche, Sigma and Stratagene.

8.1.3 Solutions

TE

10mM Tris; 50mM EDTA; adjusted to pH 8

6x Agarose gel loading buffer

0.25% bromophenol blue; 40% (w/v) sucrose in H_2O; stored at 4°C

TAE

40mM Tris-acetate ; 1mM EDTA

PBS

137mM NaCl ; 2.68mM KCL ; 10mM Na_2HPO_4 ; 1.76mM KH_2PO_4 pH 7.4

8.1.3.1 Enzymes

Restriction enzymes were purchased from New England Biolabs. *Taq* polymerase was supplied by Roche and New England Biolabs.

8.1.3.2 Plasmids

Name	Description	Reference/Source
pGEM-T	TA cloning vector, for cloning PCR products.	Promega
pUAST	Used for subloning	Brand and Perimon, 1993
TOPO	TA cloning vector for cloning PCR products	Invitrogen

8.1.3.3 Bacterial Strain

Name	Genotype and use	Reference
DH5α	deoR, *endA1*, *gyrA96*, *hsdR17* ($r_k^- m_k^+$), *supE44*, *thi-1*, *rec A1*, *relA*, *Δ(lacZYA-argF)U169*, *deoR(φ80δlacZΔM15)*, F^-, λ^-	Hanahan, 1983

8.1.3.4 *Drosophila melanogaster* strains

Name (genotype)	Reference/Source
Oregan R, wild type strain	Lab stock
W118, w-	Lab stock
w-/w-; +/+; D-Six4^{289}/TM6, Ubx	Lab stock
w-/w-; +/+; D-Six4^{289}/TM3, p[GFP]	Lab stock
w-/w-; +/+; D-Six4^{289}/TM3, Sb P[Ubx-lacZ]	Lab stock
w-/w-; +/+; P[UAS-D-Six4] D-Six4^{289}/TM3, Sb P[Ubx-lacZ]	Lab stock
w-/w-; +/+; [UAS	Lab stock

So]/[UAS-So]	
w-/w-; +/+; P[UAS-So] D-Six4^{289}/TM6 P[UBX]	Lab stock
w-/w-; +/+; P[UAS-So] D-Six4^{289}/TM3, Sb P[Ubx-lacZ]	Result of crosses
w-/w-; +/+; [UAS-Optix]/ [UAS-Optix]	Lab stock
w-/w-; +/+; [UAS-Optix] D-Six4^{289}/ TM3, Sb P[Ubx-lacZ]	Result of crosses
w-/w-; +/+; [UAS-Eya]/ [UAS-Eya]	Lab stock
w-/w-; +/+; [UAS-Eya]/ [UAS-Eya] ; D-Six4^{289}/ Tm6, Ubx	Result of crosses
w-[UAS-Six1]/ w-[UAS-Six1], +/+ ; +/+	Result of crosses
w-/w- ; [UAS-Six1]/ [UAS-Six1], +/+	Result of crosses
w-/w- ; +/+ ; [UAS-Six1]/ [UAS-Six1]	Result of crosses
w-/w- ; +/+ ; [UAS-Six1] D-Six4^{289}/ Tm3, Sb	Result of crosses

P[Ubx-lacZ]	
w- ; [UAS-Six3]/CyO ; +/+	Lab stock
w-/w-; [twist-Gal4]/[twist-Gal4], +/+	Lab stock
w-/w- ; P[twist-Gal4]/ P[twist-Gal4]; D-Six4^{289}/Tm3, Sb P[Ubx-lacZ]	Lab stock
w-; Sp/CyO,Dr,D2-3/TM6,Ubx	Lab stock
w-/w- ; +/+ ; D-Six4^{131}/Tm3, p[GFP]	Lab stock
w-; +/+; Ly/ TM3, Sb	Lab stock
w-; Pin/CyO; +/+	Lab stock

8.1.6.5 Oligonucleotides

The table shows the sequences of oligonucleotides. Underlined nucleotides are within recognition sites for restriction endonucleases.

Name	**Sequence**	**Use**
SIX5 forward primer	CAG AAT TCC AAA ACA TGG CTA CCT TGC CT	To amplify human Six5 for construct making.

SIX5 forward primer	GAG AAT TCG CCA CCA TGG CTA CCT TGC CTG CGG AG	To amplify human Six5 for construct making.
SIX5 reverse primer	CTA GAT CTA TGG AGG CAT AAC ATG TCC TAG	To amplify human Six5 for construct making.
SIX5 reverse primer	GAA GAT CTT CAC AGT TCC AAG GGC TCC TC	To amplify human Six5 for construct making.
forward	GAC TCC TAA CGC AGT CCA GGC GTG TAG CAT GTG C	To sequence human Six5.
reverse	GCA CAT GCT ACA CGC CTG GAC TGC GTT AGG AGT C	To sequence human Six5.
nested 1	TTC GGA AGC GGC CTC GGG	To sequence human Six5.
nested 2	GTC CTG CAG GAA GGC GTG GTG	To sequence human Six5.
Msix5forward	ATA TAT GCG GCC GCG CCA CCA TGG CTA CCT CGC CTG CGG AG	To amplify mouse Six5 for construct making.
Msix5reverse	GCG CGC TCT AGA TCA CAG TTC CAA GGG CTC CTC	To amplify mouse Six5 for construct making.
Msix4 forward	ATA TAT GCG GCC GCG CCA	To amplify mouse

	CCA TGT CCT CTT CCT CCC CCA C	Six4 for construct making.
Msix4 reverse	GCG CGC GGT ACC TTA TAA GTC TTG CAT GTC TTC ATC	To amplify mouse Six4 for construct making.
Msix4seqforward	CTG GGA CGG CGA GGA GAC GGT	To sequence mouse Six4.
Msix4seqreverse	GGA TGG AGG CCA CTC CCA GAG AG	To sequence mouse Six4.
Msix1 forward	ATA TAT GCG GCC GCG CCA CCA TGT CGA TGC TGC CGT CGT T	To amplify mouse Six1 for construct making.
Msix1 reverse	GCG CGC GAA TTC TTA GGA ACC CAA GTC CAC CAA ACT	To amplify mouse Six1 for construct making.
Raphie forward	CAT ATG TTC TCC ACG GAT CAG ATC CAG TGC	To amplify the Six4 Six domain and homeodomain for further crystallisation.
Ivan forward	GGA TCC TTC TCC ACG GAT CAG ATC CAG TGC	To amplify the Six4 Six domain and homeodomain for further antibody purification.

Raphie+Ivan reverse	AAG CTT CAG CAC CGA CAT GAT GTC CGG	To amplify the Six4 Six domain and homeodomain for further crystallisation and antibody purification.
sbpforward	CTG AAA AGA GCG TGC GAA AG	To make an Sbp probe.
sbpreverse	GTA ATA CGA CTC ACT ATA GGG CTG CTG CTG CAG TTC ATC AAC	To make an Sbp probe.
newsbpreverse	GTA ATA CGA CTC ACT ATA GGG CGT CTT GGC CTT CTT GAA GGA TG	To make an Sbp probe.
newsbpforward	GTA TGG ATT CGA GGG TGT GGA T	To make an Sbp probe.
obpforward	CCG AGG AGA CTA CCC CGA AC	To make an Obp probe.
obpreverse	GTA ATA CGA CTC ACT ATA GGG CGC GCA TGT GGA ATT TGA GAA ACT TCT CCG A	To make an Obp probe.

8.2 Methods

8.2.1 Manipulation of bacteria

8.2.1.1 Growth of *E.coli* cultures

E.coli cultures were grown by inoculation of bacteria from a single colony into LB or 2xTY broth and incubation for 14-16 hours at 37°C with aeration by vigorous shaking. For strains carrying ampicilin-resistant plasmids, LB was supplemented with ampicillin.

8.2.1.2 Storage of *E.coli* cultures

For long term storage, *E.coli* cultures in logarithmic phase growth were mixed with an equal volume of glycerol, placed in sterile tubes and kept at -70°C. To grow bacteria from frozen culture, a small portion was removed using a sterile loop and streaked on an L-agar plate, with ampicillin if required.

For short term storage up to six weeks, bacteria were streaked onto agar plates which were incubated at 37°C 14-16 hours for colony growth then kept at 4°C.

8.2.1.3 Transformation of bacteria

Transformation of *E.coli* by purified plasmid DNA or ligation products was carried out by electroporation according to Heery and Dunican (1989). Cultures were grown to early stationary phase in 2xTY medium and cells harvested by centrifugation at 9000xG for 10 minutes at 4°C. Cells were washed by resuspension in a volume of ice-cold dH_2O equal to that of the original culture and again collected by centrifugation. This wash was repeated twice more and the cells were resuspended in an equal volume of ice-cold dH_2O. 40µl of cells were mixed with 1µl DNA solution

and transferred to an electroporation cuvette (0.2cm, Invitrogen). A single pulse at 2.5kV, 15μF, 200Ω was applied. 1ml SOC buffer was added immediately and the mixture transferred to a culture tube. Cells were then incubated at 37°C with shaking for 20 minutes. Several dilutions in SOC buffer were made and plated on L-agar with ampicillin. For selection for inactivation of β galactosidase expression 100μl of 100mM IPTG and 20μl 50mg/ml X-gal were spread onto the plates which were then incubated for 30 minutes at 37°C for absorption prior to use.

8.2.2 *In vitro* manipulation of DNA

8.2.2.1 Small scale preparation of plasmid DNA

Small scale preparation of plasmid DNA from *E.coli* cultures was carried out using the Wizard® *Plus* SV minipreps DNA purification system (promega) according to the manufacturer's instructions. This method involves alkaline lysis of bacteria followed by a brief treatment with alkaline protease to inactivate endonucleasese released on cell lysis. Plasmid DNA is then purified by binding to a column, washing in a 60% ethanol solution to remove impurities and finally elution in dH_2O.

8.2.2.2 Large scale preparation of plasmid DNA

Preparation of up to 100μg of plasmid DNA from *E.coli* cultures was carried out using the Qiagen plasmid midi kit (Qiagen GmbH and Qiagen Inc) according to the manufacturer's directions. This method is similar to the miniprep method described previously. Alkaline lysis of *E.coli* is followed by binding of plasmid DNA to an anion exchange resin under low salt and pH conditions. The resin is washed in a

medium salt buffer, and the DNA eluted by high salt. Finally the DNA is concentrated by isopropanol precipitation.

8.2.2.3 Large scale preparation of plasmid DNA for injections

Liquid bacterial cultures were transferred to 50ml Falcon tubes and centrifuged at 1000rpm for 20 minutes at 4°C. The pellets were drained thoroughly and resuspended carefully using a pastette in 2ml of the solution 1 (50mM Glucose, 25mM Tris pH 8, 10mM EDTA, 5mg/ml lysozyme, prepared just before use) per 50ml of culture and left at room temperature for 10 minutes. 4ml of solution 2 (0.2 M NaOH, 1% SDS- prepared just before use) was added and mixed thoroughly but not vigorously. The viscous mixture was incubated on ice for 10 minutes with regular gentle agitation. 3ml of solution 3 (3M KOAc, 1.3M HCOOH) was added with immediate, thorough mixing and placed on ice for 15 minutes. The mixture was centrifuged at 4500rpm for 15 minutes. The clear supernatant was transferred to a clean tube avoiding transfer of any precipitate. 0.6 (v/v) of 100% isopropanol was added and the solution was mixed and incubated at RT for 5 minutes. The tube was then centrifuged at 4,500rpm for 10 minutes. The supernatant was discarded and the pellet rinsed with 2ml of 70% ethanol. The inner walls of the tubes were wiped clean and the still wet pellet dissolved in 1ml TE. The DNA solution was transferred to eppendorfs and placed on ice for 5-10 minutes. An equal volume of cold 5M LiCl (stored at -20°C) was added and the tubes were incubated on ice for 5 minutes, followed by centrifugation at 14,000rpm for 5 minutes. The supernatant was transferred to clean eppendorf tubes (on ice) and an equal volume of isopropanol was added. The tubes were incubated on ice for 10 minutes and then centrifuged at

14,000rpm for 5 minutes. The supernatant was discarded and the pellets air-dried at RT. The pellets were then resuspended in a total of 300µl TE.

To remove RNA, 1µl DNAse-free RNAse (10mg/ml stock) was added and the mixture incubated at 37°C for 30 minutes. The mixture was then transferred to ice and an equal volume of PEG/NaCl (15% PEG, 1.6M NaCl) was added. This mixture was then incubated on ice for 5 minutes before centrifiguation at 14,000rpm for 5 minutes. The supernatant was discarded and the pellet was resuspended in 300µl TE. The plasmid DNA was then purified by PhOH/$CHCl_3$ extraction. The DNA was precipitated by addition of 0.05 (v/v) 3M NaOAc (pH 5.2-5.6) and 2 (v/v) 100% ethanol. This was thoroughly mixed and incubated at -20°C overnight. The tubes were then centrifuged at 14,000rpm for 5 minutes. The pellets were then washed with 70% ethanol, air-dried and resuspended in 300µl ddH_2O.

8.2.2.4 Precipitation of DNA using ethanol

DNA in solution was precipitated by the addition of 1/9 volume 3M sodium acetate pH 5.2 followed by 3 volumes absolute ethanol. After mixing, the solution was incubated 20 minutes on ice and DNA recovered by centrifugation at 13,000rpm for 10 minutes in a Biofuge 13 microcentrifuge (Heraeus). Following removal of the supernatatnt the pellet was washed with 70% ethanol and dried for 10 minutes at room temperature. DNA was dissolved in dH_2O or TE.

8.2.2.5 Quantification of DNA

DNA concentrations were estimated by measurement of absorption at 260nm using a lambda UV/VIS spectrophotometer (Perkin Elmer). Absorption measurements were converted to DNA concentrations using an extinction coefficient of 50μg/ml for double-stranded DNA and 33μg/ml for single-stranded DNA.

8.2.2.6 Cleavage of DNA by restriction endonucleases

DNA cleavage was carried out using enzymes and buffers supplied by Boehringer Mannheim and New England Biolabs under the conditions recommended by the manufacturers. Digests of 0.1 to 20μg DNA were carried out in 20-100μl of the appropriate 1xreaction buffer for 1-12 hours at 37°C.

8.2.2.7 Agarose gel electrophoresis

Electrophoresis of DNA was carried out in 0.7-2% MP agarose (Boehringer Mannheim) in TAE containing 0.5mg/ml ethidium bromide. Prior to loading, DNA samples were mixed with 1/6 volume 6xagarose gel loading buffer. A potential difference of 1-10V per cm gel was used to separate DNA fragments. Following electrophoresis DNA was visualised and photographed on a UV transiluminator.

8.2.2.8 Purification of DNA fragments from agarose

Gel slices containing DNA fragments separated by agarose gel electrophoresis were purified using the QIAquick gel extraction kit (Qiagen) according to the manufacturer's instructions. This involves binding of DNA to a silica gel membrane

at low pH in the presence of chaotropic salt, followed by washing in a buffer containing ethanol and low salt elution in 10mM Tris.CL pH8.5.

8.2.2.9 Ligation of DNA fragments 1

Ligation of PCR products into pGEM®-T was carried out according to the instructions of the manufacturer of the pGEM®-T vector system cloning kit (Promega).

8.2.2.10 Ligation of DNA fragments 2

In order to maximise the ligation between vector and insert fragments, a standard formula was used to predict the best fragment vector ratios.

[vector (ng) x fragment size (bp)/ vector size (bp)] x 3 = ng of insert needed

T4 DNA ligase (NEB) was used according to the manufacturer's instructions. Ligations were performed at 16°C overnight.

8.2.2.11 Random primed labelling

Probes for hybridisation were prepared by random-primed labelling using the T7 quickprime kit (Pharmacia) according to the instructions of the manufacturer. 10-100ng of PCR product were added to a reagent mix containing random oligonucleotides, buffer and dNTPs. T7 DNA polymerase was added and the mixture incubated for 2 hours at 37°C.

8.2.2.12 Sequencing of double-stranded plasmid DNA

DNA reactions are sent off for sequencing to the Ashworth Sequencing Service in King's Buildings, Edinburgh.

500ng of DNA and 3.2pmole/µl are added to 3.68µl ddH_2O, 2µl 5XSequencing Buffer and 2µl Big Dye and are subjected to 25 cycles of 95°C for 30 seconds, 50°C for 20 seconds and 60°C for 4 minutes.

The samples are then cleaned up using Edge Biosystems Performa DTR V3 plates which remove dNTP's, salts, label from probes and other low molecular weight material. The samples are transferred to plates and dried down in a vacuum concentrator. 10µl of Hi-Di Formamide is added to each well and the plate put on a brief heat cycle of 95°C for 2 minutes and cooled back to 4°C. The plate is then put on the sequencer machine (3730 DNA Analyzer) and the samples are run on a 50cm array with POP-7 polymer.

8.2.2.13 Polymerase chain reaction

PCRs were carried out in 50µl of the appropriate PCR buffer with 0.5µM each primer, 0.25mM each dNTP (Boehringer Mannheim), and 1 unit *Taq* DNA polymerase (Promega). Reactions were incubated at 95°C for 10 minutes followed by 30 cycles of 1min at 95°C, 1 minute at the annealing temperature, (generally 55°C unless otherwise specified) and 1minute at 72°C. Finally reactions were incubated at 72°C for 10 minutes.

8.2.2.14 PCR product processing

PCR products were cloned using the pGEM®-T vector system I kit (Promega). This utilises a pre-cut plasmid vector (pGEM®-T) having a single unpaired

deoxythymidine nucleotide at each 3' end This provides compatible overhangs for ligation to PCR products as thermostable polymersases add an unpaired deoxyadenosine to the 5' end during synthesis (Clark, 1988). Following ligation and transformation, colonies with plasmids containing insertions are detected by blue-white selection.

8.2.3 Manipulation of *Drosophila melanogaster* flies and tissues

8.2.3.1 Maintenance of *Drosophila* stocks

Drosophila melanogaster strains were maintained at 25°C on Dundee fly food. To maintain the reactivity of stocks, only flies up to seven days old were used for breeding.

8.2.3.2 Collection of *Drosophila* developmental stages

Embryos were collected on egg collection medium in Petri dishes placed on the bottom of fly cages. Plates were spread with yeast to provide food for the flies. After allowing females to lay eggs for the appropriate length of time, plates were collected and embryos were washed off onto nylon mesh with distilled water. Embryos were then thoroughly washed with distilled water, and collected in Eppendorf tubes.

Adult flies were anaesthetised (with carbon dioxide), sexed and collected in Eppendorf tubes.

8.2.3.3 Fixation of embryos for immunohistochemistry

Embryos were collected on grape juice plates with a globule of yeast paste (20% glucose) as a nutrient source. The grape juice plates were then aged for the appropriate length of time at appropriate temperatures (see table below). The embryos were removed using ddH2O and a paintbrush, and pipetted into a fine sieve. Embryos were washed to remove yeast and dechorionated in 50% fresh bleach for 4 minutes, then thoroughly washed to remove bleach. The embryos were then transferred into a scintillation vial and fixed for 20 minutes with agitation in 1.25ml formaldehyde (37%), 3.75ml PBS (8g NaCl, 0.2g KCLl, 1.44g Na2HPO4, 0.24g KH2PO4 for 1 litre, adjusted to pH 7.4) and 5 ml n-Heptane (Sigma). The bottom phase of formaldehyde was removed and 10ml of methanol was added. The scintillation vial was then shaken for 30 seconds to devitellinise the embryos. Embryos were allowed to settle to the bottom of the vial and then transferred to an eppendorf. The embryos were then washed with methanol to remove residual heptane, and then washed 4 times with PBST. This was followed by the standard wash procedure.

Embryos were blocked for at least two hours in 2% bovine serum albumin (BSA) solution (Sigma) in PBST at room temperature on a rotating wheel. Primary antibody, in PBST at the appropriate concentration with 0.5% (v/v) BSA, 0.05% (v/v) Normal Goat Serum (NGS, Jackson labs) was added and samples were incubated at 4°C overnight. The primary antibodies were then rinsed with the standard wash procedure. The secondary antibody (fluorochrome conjugate) was added in PBST to a concentration of 1:1000 for 2 hours at room temperature. The samples were rinsed with the standard wash procedure, then mounted in Vectashield

(Vector labs) on microscope slides sealed with a cover slip and nail varnish. Slides were stored in the dark at 4°C. Confocal images were taken on a Zeiss LSM5 Pascal confocal microscope.

8.2.3.4 Preparation of *Drosophila* genomic DNA

20 flies were frozen for 5 minutes at -70°C then resuspended in 400µl lysis buffer. The flies were homogenised using a hand-held Pellet-pestle® motor homogeniser (Kontes). Following incubation at 70°C for 30 minutes, 56µl 8M potassium acetate was added. Samples were incubated 30 minutes on ice. To remove insoluble material, samples were centrifuged at 4°C for 15 minutes at full speed in a microcentrifuge. The supernatant was removed and the centrifugation repeated. The final supernatant was added to 200µl isopropanol and cooled at -70°C for 10 minutes for precipitation of DNA. DNA was recovererd by centrifugation, washed with 70%, dried and resuspended in 40µl TE.

8.2.3.5 Production of transformant fly lines by microinjection

Constructs containing the Six gene of interest in a pUASt vector (P element vector) were injected into Δ2-3 flies. The Δ 2-3 is the source of transposase for the attenuated P element vector. DNA is introduced into pr-cellular blastoderm embryos by injection and integrated into the genome by random transposition events. DNA for each construct was prepared using the method described above.

Cages of flies were set up and the grape-juice agar plate with yeast paste changed regularly to encourage laying. Plates were collected every hour and the embryos used for injection. The injection procedure was carried out at 18°C.

Embryos were dechorionated for 4 minutes in 50% bleach and then rinsed in H_2O. Embryos were lined up under a microscope along the edge of a piece of agar in one orientation. They were then transferred to a cover slip coated with a film of glue. The cover slip was attached to a microscope slide using a drop of oil and placed at 18°C for 20 minutes. It was then transferred to silica beads 18°C for 10 minutes to allow dehydration. Embryos were then covered with series 700 halocarbon oil and injected with the construct of interest. Injected embryos were then covered in series 95 halocarbon oil, left at 18°C for two days and then allowed to develop at 25°C. Adult flies were crossed with white eyed flies (w118) and transformants screened for on the basis of eye colour.

8.2.4 Computer Programs

cDNA sequences of the six genes were acquired from Flybase sequences (http:flybase.bio.indiana.edu/). These were used to design primers for the making of constructs.

References

(2006). Insights into social insects from the genome of the honeybee Apis mellifera. *Nature* **443,** 931-49.

Adams, M. D., Celniker, S. E., Holt, R. A., Evans, C. A., Gocayne, J. D., Amanatides, P. G., Scherer, S. E., Li, P. W., Hoskins, R. A., Galle, R. F., George, R. A., Lewis, S. E., Richards, S., Ashburner, M., Henderson, S. N., Sutton, G. G., Wortman, J. R., Yandell, M. D., Zhang, Q., Chen, L. X., Brandon, R. C., Rogers, Y. H., Blazej, R. G., Champe, M., Pfeiffer, B. D., Wan, K. H., Doyle, C., Baxter, E. G., Helt, G., Nelson, C. R., Gabor, G. L., Abril, J. F., Agbayani, A., An, H. J., Andrews-Pfannkoch, C., Baldwin, D., Ballew, R. M., Basu, A., Baxendale, J., Bayraktaroglu, L., Beasley, E. M., Beeson, K. Y., Benos, P. V., Berman, B. P., Bhandari, D., Bolshakov, S., Borkova, D., Botchan, M. R., Bouck, J., Brokstein, P., Brottier, P., Burtis, K. C., Busam, D. A., Butler, H., Cadieu, E., Center, A., Chandra, I., Cherry, J. M., Cawley, S., Dahlke, C., Davenport, L. B., Davies, P., de Pablos, B., Delcher, A., Deng, Z., Mays, A. D., Dew, I., Dietz, S. M., Dodson, K., Doup, L. E., Downes, M., Dugan-Rocha, S., Dunkov, B. C., Dunn, P., Durbin, K. J., Evangelista, C. C., Ferraz, C., Ferriera, S., Fleischmann, W., Fosler, C., Gabrielian, A. E., Garg, N. S., Gelbart, W. M., Glasser, K., Glodek, A., Gong, F., Gorrell, J. H., Gu, Z., Guan, P., Harris, M., Harris, N. L., Harvey, D., Heiman, T. J., Hernandez, J. R., Houck, J., Hostin, D., Houston, K. A., Howland, T. J., Wei, M. H., Ibegwam, C., et al. (2000). The genome sequence of Drosophila melanogaster. *Science* **287,** 2185-95.

Aijaz, S., Clark, B. J., Williamson, K., van Heyningen, V., Morrison, D., Fitzpatrick, D., Collin, R., Ragge, N., Christoforou, A., Brown, A., and Hanson, I. (2004). Absence of SIX6 mutations in microphthalmia, anophthalmia, and coloboma. *Invest Ophthalmol Vis Sci* **45,** 3871-6.

Ando, Z., Sato, S., Ikeda, K., and Kawakami, K. (2005). Slc12a2 is a direct target of two closely related homeobox proteins, Six1 and Six4. *Febs J* **272,** 3026-41.

Appolloni, I., Calzolari, F., Corte, G., Perris, R., and Malatesta, P. (2007). Six3 Controls the Neural Progenitor Status in the Murine CNS. *Cereb Cortex.*

Arnosti, D. N. (2003). Analysis and function of transcriptional regulatory elements: insights from Drosophila. *Annu Rev Entomol* **48,** 579-602.

Artero, R., Furlong, E. E., Beckett, K., Scott, M. P., and Baylies, M. (2003). Notch and Ras signaling pathway effector genes expressed in fusion competent and founder cells during Drosophila myogenesis. *Development* **130,** 6257-72.

Azpiazu, N., and Frasch, M. (1993). tinman and bagpipe: two homeo box genes that determine cell fates in the dorsal mesoderm of Drosophila. *Genes Dev* **7,** 1325-40.

Bate, M., and Rushton, E. (1993). Myogenesis and muscle patterning in Drosophila. *C R Acad Sci III* **316,** 1047-61.

Bate, M., Rushton, E., and Currie, D. A. (1991). Cells with persistent twist expression are the embryonic precursors of adult muscles in Drosophila. *Development* **113,** 79-89.

Baylies, M. K., and Bate, M. (1996). twist: a myogenic switch in Drosophila. *Science* **272,** 1481-4.

Bebenek, I. G., Gates, R. D., Morris, J., Hartenstein, V., and Jacobs, D. K. (2004). sine oculis in basal Metazoa. *Dev Genes Evol* **214,** 342-51.

Bodmer, R. (1993). The gene tinman is required for specification of the heart and visceral muscles in Drosophila. *Development* **118,** 719-29.

Boucher, C. A., Winchester, C. L., Hamilton, G. M., Winter, A. D., Johnson, K. J., and Bailey, M. E. (2000). Structure, mapping and expression of the human gene encoding the homeodomain protein, SIX2. *Gene* **247,** 145-51.

Bour, B. A., O'Brien, M. A., Lockwood, W. L., Goldstein, E. S., Bodmer, R., Taghert, P. H., Abmayr, S. M., and Nguyen, H. T. (1995). Drosophila MEF2, a transcription factor that is essential for myogenesis. *Genes Dev* **9,** 730-41.

Bourgouin, C., Lundgren, S. E., and Thomas, J. B. (1992). Apterous is a Drosophila LIM domain gene required for the development of a subset of embryonic muscles. *Neuron* **9,** 549-61.

Bovolenta, P., Mallamaci, A., Puelles, L., and Boncinelli, E. (1998). Expression pattern of cSix3, a member of the Six/sine oculis family of transcription factors. *Mech Dev* **70,** 201-3.

Boyle, M., Bonini, N., and DiNardo, S. (1997). Expression and function of clift in the development of somatic gonadal precursors within the Drosophila mesoderm. *Development* **124,** 971-82.

Boyle, M., and DiNardo, S. (1995). Specification, migration and assembly of the somatic cells of the Drosophila gonad. *Development* **121,** 1815-25.

Brand, A. H., and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118,** 401-15.

Bricaud, O., and Collazo, A. (2006). The transcription factor six1 inhibits neuronal and promotes hair cell fate in the developing zebrafish (Danio rerio) inner ear. *J Neurosci* **26,** 10438-51.

Brodbeck, S., Besenbeck, B., and Englert, C. (2004). The transcription factor Six2 activates expression of the Gdnf gene as well as its own promoter. *Mech Dev* **121,** 1211-22.

Brodbeck, S., and Englert, C. (2004). Genetic determination of nephrogenesis: the Pax/Eya/Six gene network. *Pediatr Nephrol* **19,** 249-55.

Brody, T. (1999). The Interactive Fly: gene networks, development and the Internet. *Trends Genet* **15,** 333-4.

Bryant, P. J., Adler, P. N., Duranceau, C., Fain, M. J., Glenn, S., Hsei, B., James, A. A., Littlefield, C. L., Reinhardt, C. A., Strub, S., and Schneiderman, H. A. (1978). Regulative interactions between cells from different imaginal disks of Drosophila melanogaster. *Science* **201,** 928-30.

Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. *Nucleic Acids Res* **15,** 1353-61.

Cheyette, B. N., Green, P. J., Martin, K., Garren, H., Hartenstein, V., and Zipursky, S. L. (1994). The Drosophila sine oculis locus encodes a homeodomain-containing protein required for the development of the entire visual system. *Neuron* **12,** 977-96.

Chi, Y. I. (2005). Homeodomain revisited: a lesson from disease-causing mutations. *Hum Genet* **116,** 433-44.

Clark, I. B., Boyd, J., Hamilton, G., Finnegan, D. J., and Jarman, A. P. (2006). D-six4 plays a key role in patterning cell identities deriving from the Drosophila mesoderm. *Dev Biol* **294,** 220-31.

Clark, I. B., Jarman, A. P., and Finnegan, D. J. (2007). Live imaging of Drosophila gonad formation reveals roles for Six4 in regulating germline and somatic cell migration. *BMC Dev Biol* **7,** 52.

Clark, S. W., Fee, B. E., and Cleveland, J. L. (2002). Misexpression of the eyes absent family triggers the apoptotic program. *J Biol Chem* **277,** 3560-7.

Conte, I., Morcillo, J., and Bovolenta, P. (2005). Comparative analysis of Six 3 and Six 6 distribution in the developing and adult mouse brain. *Dev Dyn* **234,** 718-25.

Crosby, M. A., Goodman, J. L., Strelets, V. B., Zhang, P., and Gelbart, W. M. (2007). FlyBase: genomes by the dozen. *Nucleic Acids Res* **35,** D486-91.

DeFalco, T., Le Bras, S., and Van Doren, M. (2004). Abdominal-B is essential for proper sexually dimorphic development of the Drosophila gonad. *Mech Dev* **121,** 1323-33.

DeFalco, T. J., Verney, G., Jenkins, A. B., McCaffery, J. M., Russell, S., and Van Doren, M. (2003). Sex-specific apoptosis regulates sexual dimorphism in the Drosophila embryonic gonad. *Dev Cell* **5,** 205-16.

Dohrmann, C., Azpiazu, N., and Frasch, M. (1990). A new Drosophila homeo box gene is expressed in mesodermal precursor cells of distinct muscles during embryogenesis. *Genes Dev* **4,** 2098-111.

Dozier, C., Kagoshima, H., Niklaus, G., Cassata, G., and Burglin, T. R. (2001). The Caenorhabditis elegans Six/sine oculis class homeobox gene ceh-32 is required for head morphogenesis. *Dev Biol* **236,** 289-303.

Duboule, D. (1994). "Guidebook to the homeobox genes." Sambrook & Tooze Publication at Oxford University Press, Oxford.

Edgley, M. L., Baillie, D. L., Riddle, D. L., and Rose, A. M. (1995). Genetic balancers. *Methods Cell Biol* **48,** 147-84.

Esteve, P., and Bovolenta, P. (1999). cSix4, a member of the six gene family of transcription factors, is expressed during placode and somite development. *Mech Dev* **85,** 161-5.

Evans, S. M. (1999). Vertebrate tinman homologues and cardiac differentiation. *Semin Cell Dev Biol* **10,** 73-83.

Fischbach, K. F., and Technau, G. (1984). Cell degeneration in the developing optic lobes of the sine oculis and small-optic-lobes mutants of Drosophila melanogaster. *Dev Biol* **104,** 219-39.

Ford, H. L., Landesman-Bollag, E., Dacwag, C. S., Stukenberg, P. T., Pardee, A. B., and Seldin, D. C. (2000). Cell cycle-regulated phosphorylation of the human SIX1 homeodomain protein. *J Biol Chem* **275,** 22245-54.

Fougerousse, F., Durand, M., Lopez, S., Suel, L., Demignon, J., Thornton, C., Ozaki, H., Kawakami, K., Barbet, P., Beckmann, J. S., and Maire, P. (2002). Six and Eya expression during human somitogenesis and MyoD gene family activation. *J Muscle Res Cell Motil* **23,** 255-64.

Frasch, M. (1995). Induction of visceral and cardiac mesoderm by ectodermal Dpp in the early Drosophila embryo. *Nature* **374,** 464-7.

Furlong, E. E., Andersen, E. C., Null, B., White, K. P., and Scott, M. P. (2001). Patterns of gene expression during Drosophila mesoderm development. *Science* **293,** 1629-33.

Furuya, M., Qadota, H., Chisholm, A. D., and Sugimoto, A. (2005). The C. elegans eyes absent ortholog EYA-1 is required for tissue differentiation and plays partially redundant roles with PAX-6. *Dev Biol* **286,** 452-63.

Gilboa, L., and Lehmann, R. (2006). Soma-germline interactions coordinate homeostasis and growth in the Drosophila gonad. *Nature* **443,** 97-100.

Giordani, J., Bajard, L., Demignon, J., Daubas, P., Buckingham, M., and Maire, P. (2007). Six proteins regulate the activation of Myf5 expression in embryonic mouse limbs. *Proc Natl Acad Sci U S A* **104,** 11310-5.

Goudreau, G., Petrou, P., Reneker, L. W., Graw, J., Loster, J., and Gruss, P. (2002). Mutually regulated expression of Pax6 and Six3 and its implications for the Pax6 haploinsufficient lens phenotype. *Proc Natl Acad Sci U S A* **99,** 8719-24.

Greig, S., and Akam, M. (1993). Homeotic genes autonomously specify one aspect of pattern in the Drosophila mesoderm. *Nature* **362,** 630-2.

Grifone, R., Demignon, J., Giordani, J., Niro, C., Souil, E., Bertin, F., Laclef, C., Xu, P. X., and Maire, P. (2007). Eya1 and Eya2 proteins are required for hypaxial somitic myogenesis in the mouse embryo. *Dev Biol* **302,** 602-16.

Grifone, R., Demignon, J., Houbron, C., Souil, E., Niro, C., Seller, M. J., Hamard, G., and Maire, P. (2005). Six1 and Six4 homeoproteins are required for Pax3 and Mrf expression during myogenesis in the mouse embryo. *Development* **132,** 2235-49.

Grifone, R., Laclef, C., Spitz, F., Lopez, S., Demignon, J., Guidotti, J. E., Kawakami, K., Xu, P. X., Kelly, R., Petrof, B. J., Daegelen, D., Concordet, J. P., and Maire, P. (2004). Six1 and Eya1 expression can reprogram adult muscle from the slow-twitch phenotype into the fast-twitch phenotype. *Mol Cell Biol* **24,** 6253-67.

Gustafson, K., and Boulianne, G. L. (1996). Distinct expression patterns detected within individual tissues by the GAL4 enhancer trap technique. *Genome* **39,** 174-82.

Halder, G., Callaerts, P., Flister, S., Walldorf, U., Kloter, U., and Gehring, W. J. (1998). Eyeless initiates the expression of both sine oculis and eyes absent during Drosophila compound eye development. *Development* **125,** 2181-91.

Hanson, I. M. (2001). Mammalian homologues of the Drosophila eye specification genes. *Semin Cell Dev Biol* **12,** 475-84.

Hauser, F., Cazzamali, G., Williamson, M., Blenau, W., and Grimmelikhuijzen, C. J. (2006). A review of neurohormone GPCRs present in the fruitfly Drosophila melanogaster and the honey bee Apis mellifera. *Prog Neurobiol* **80,** 1-19.

Heanue, T. A., Reshef, R., Davis, R. J., Mardon, G., Oliver, G., Tomarev, S., Lassar, A. B., and Tabin, C. J. (1999). Synergistic regulation of vertebrate muscle development by Dach2, Eya2, and Six1, homologs of genes required for Drosophila eye formation. *Genes Dev* **13,** 3231-43.

Holt, R. A., Subramanian, G. M., Halpern, A., Sutton, G. G., Charlab, R., Nusskern, D. R., Wincker, P., Clark, A. G., Ribeiro, J. M., Wides, R., Salzberg, S. L., Loftus, B., Yandell, M., Majoros, W. H., Rusch, D. B., Lai, Z., Kraft, C. L., Abril, J. F., Anthouard, V., Arensburger, P., Atkinson, P. W., Baden, H., de Berardinis, V., Baldwin, D., Benes, V., Biedler, J., Blass, C., Bolanos, R., Boscus, D., Barnstead, M., Cai, S., Center, A., Chaturverdi, K., Christophides, G. K., Chrystal, M. A., Clamp, M., Cravchik, A., Curwen, V., Dana, A., Delcher, A., Dew, I., Evans, C. A., Flanigan, M., Grundschober-Freimoser, A., Friedli, L., Gu, Z., Guan, P., Guigo, R., Hillenmeyer, M. E., Hladun, S. L., Hogan, J. R., Hong, Y. S., Hoover, J., Jaillon, O., Ke, Z., Kodira, C., Kokoza, E., Koutsos, A., Letunic, I., Levitsky, A., Liang, Y., Lin, J. J., Lobo, N. F., Lopez, J. R., Malek, J. A., McIntosh, T. C., Meister, S., Miller, J., Mobarry, C., Mongin, E., Murphy, S. D., O'Brochta, D. A., Pfannkoch, C., Qi, R., Regier, M. A., Remington, K., Shao, H., Sharakhova, M. V., Sitter, C. D., Shetty, J., Smith, T. J., Strong, R., Sun, J., Thomasova, D., Ton, L. Q., Topalis, P., Tu, Z., Unger, M. F., Walenz, B., Wang, A., Wang, J., Wang, M., Wang, X., Woodford, K. J., Wortman, J. R., Wu, M., Yao, A., Zdobnov, E.

M., Zhang, H., Zhao, Q., et al. (2002). The genome sequence of the malaria mosquito Anopheles gambiae. *Science* **298,** 129-49.

Hoshiyama, D., Iwabe, N., and Miyata, T. (2007). Evolution of the gene families forming the Pax/Six regulatory network: isolation of genes from primitive animals and molecular phylogenetic analyses. *FEBS Lett* **581,** 1639-43.

Hoskins, B. E., Cramer, C. H., Silvius, D., Zou, D., Raymond, R. M., Orten, D. J., Kimberling, W. J., Smith, R. J., Weil, D., Petit, C., Otto, E. A., Xu, P. X., and Hildebrandt, F. (2007). Transcription factor SIX5 is mutated in patients with branchio-oto-renal syndrome. *Am J Hum Genet* **80,** 800-4.

Hu, S., Mamedova, A., and Hegde, R. S. (2008). DNA-Binding and Regulation Mechanisms of the SIX Family of Retinal Determination Proteins. *Biochemistry* **47,** 3586-94.

Hubbard, T. J., Aken, B. L., Beal, K., Ballester, B., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cunningham, F., Cutts, T., Down, T., Dyer, S. C., Fitzgerald, S., Fernandez-Banet, J., Graf, S., Haider, S., Hammond, M., Herrero, J., Holland, R., Howe, K., Howe, K., Johnson, N., Kahari, A., Keefe, D., Kokocinski, F., Kulesha, E., Lawson, D., Longden, I., Melsopp, C., Megy, K., Meidl, P., Ouverdin, B., Parker, A., Prlic, A., Rice, S., Rios, D., Schuster, M., Sealy, I., Severin, J., Slater, G., Smedley, D., Spudich, G., Trevanion, S., Vilella, A., Vogel, J., White, S., Wood, M., Cox, T., Curwen, V., Durbin, R., Fernandez-Suarez, X. M., Flicek, P., Kasprzyk, A., Proctor, G., Searle, S., Smith, J., Ureta-Vidal, A., and Birney, E. (2007). Ensembl 2007. *Nucleic Acids Res* **35,** D610-7.

Ikeda, K., Watanabe, Y., Ohto, H., and Kawakami, K. (2002). Molecular interaction and synergistic activation of a promoter by Six, Eya, and Dach proteins mediated through CREB binding protein. *Mol Cell Biol* **22,** 6759-66.

Inukai, A., Doyu, M., Kato, T., Liang, Y., Kuru, S., Yamamoto, M., Kobayashi, Y., and Sobue, G. (2000). Reduced expression of DMAHP/SIX5 gene in myotonic dystrophy muscle. *Muscle Nerve* **23,** 1421-6.

Ip, Y. T., Park, R. E., Kosman, D., Yazdanbakhsh, K., and Levine, M. (1992). dorsal-twist interactions establish snail expression in the presumptive mesoderm of the Drosophila embryo. *Genes Dev* **6,** 1518-30.

Jagla, K., Frasch, M., Jagla, T., Dretzen, G., Bellard, F., and Bellard, M. (1997). ladybird, a new component of the cardiogenic pathway in Drosophila required for diversification of heart precursors. *Development* **124,** 3471-9.

Jagla, K., Jagla, T., Heitzler, P., Dretzen, G., Bellard, F., and Bellard, M. (1997). ladybird, a tandem of homeobox genes that maintain late wingless expression in terminal and dorsal epidermis of the Drosophila embryo. *Development* **124,** 91-100.

Jean, D., Bernier, G., and Gruss, P. (1999). Six6 (Optx2) is a novel murine Six3-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. *Mech Dev* **84,** 31-40.

Johnson, K. J., Boucher, C., King, S. K., Winchester, C. L., Bailey, M. E., Hamilton, G. M., and Carey, N. (1996). Is myotonic dystrophy a single-gene disorder? *Biochem Soc Trans* **24,** 510-3.

Kablar, B. (2004). MyoD-lacZ transgenes are early markers in the neural retina, but MyoD function appears to be inhibited in the developing retinal cells. *Int J Dev Neurosci* **22,** 215-24.

Kawakami, K., Ohto, H., Ikeda, K., and Roeder, R. G. (1996). Structure, function and expression of a murine homeobox protein AREC3, a homologue of Drosophila

sine oculis gene product, and implication in development. *Nucleic Acids Res* **24,** 303-10.

Kawakami, K., Ohto, H., Takizawa, T., and Saito, T. (1996). Identification and expression of six family genes in mouse retina. *FEBS Lett* **393,** 259-63.

Kawakami, K., Sato, S., Ozaki, H., and Ikeda, K. (2000). Six family genes--structure and function as transcription factors and their roles in development. *Bioessays* **22,** 616-26.

Kenyon, K. L., Li, D. J., Clouser, C., Tran, S., and Pignoni, F. (2005). Fly SIX-type homeodomain proteins Sine oculis and Optix partner with different cofactors during eye development. *Dev Dyn* **234,** 497-504.

Kenyon, K. L., Yang-Zhou, D., Cai, C. Q., Tran, S., Clouser, C., Decene, G., Ranade, S., and Pignoni, F. (2005). Partner specificity is essential for proper function of the SIX-type homeodomain proteins Sine oculis and Optix during fly eye development. *Dev Biol* **286,** 158-68.

Kim, K. H., Lee, Y. S., Jeon, H. K., Park, J. K., Kim, C. B., and Eom, K. S. (2007). Hox genes from the tapeworm Taenia asiatica (Platyhelminthes: Cestoda). *Biochem Genet* **45,** 335-43.

Kirby, R. J., Hamilton, G. M., Finnegan, D. J., Johnson, K. J., and Jarman, A. P. (2001). Drosophila homolog of the myotonic dystrophy-associated gene, SIX5, is required for muscle and gonad development. *Curr Biol* **11,** 1044-9.

Kmita-Cunisse, M., Loosli, F., Bierne, J., and Gehring, W. J. (1998). Homeobox genes in the ribbonworm Lineus sanguineus: evolutionary implications. *Proc Natl Acad Sci U S A* **95,** 3030-5.

Kobayashi, M. (2000). [Six family genes as key factors in the sensory organ formation]. *Tanpakushitsu Kakusan Koso* **45,** 2775-81.

Kobayashi, M., Nishikawa, K., Suzuki, T., and Yamamoto, M. (2001). The homeobox protein Six3 interacts with the Groucho corepressor and acts as a transcriptional repressor in eye and forebrain formation. *Dev Biol* **232,** 315-26.

Kobayashi, M., Osanai, H., Kawakami, K., and Yamamoto, M. (2000). Expression of three zebrafish Six4 genes in the cranial sensory placodes and the developing somites. *Mech Dev* **98,** 151-5.

Kochhar, A., Fischer, S. M., Kimberling, W. J., and Smith, R. J. (2007). Branchio-oto-renal syndrome. *Am J Med Genet A* **143,** 1671-8.

Konishi, Y., Ikeda, K., Iwakura, Y., and Kawakami, K. (2006). Six1 and Six4 promote survival of sensory neurons during early trigeminal gangliogenesis. *Brain Res* **1116,** 93-102.

Kosman, D., Ip, Y. T., Levine, M., and Arora, K. (1991). Establishment of the mesoderm-neuroectoderm boundary in the Drosophila embryo. *Science* **254,** 118-22.

Kozmik, Z., Holland, N. D., Kreslova, J., Oliveri, D., Schubert, M., Jonasova, K., Holland, L. Z., Pestarino, M., Benes, V., and Candiani, S. (2007). "Pax-Six-Eya-Dach network during amphioxus development: conservation in vitro but context specificity in vivo."

Kumar, J. P., and Moses, K. (2001). Expression of evolutionarily conserved eye specification genes during Drosophila embryogenesis. *Dev Genes Evol* **211,** 406-14.

Laclef, C., Hamard, G., Demignon, J., Souil, E., Houbron, C., and Maire, P. (2003). Altered myogenesis in Six1-deficient mice. *Development* **130,** 2239-52.

Lagutin, O., Zhu, C. C., Furuta, Y., Rowitch, D. H., McMahon, A. P., and Oliver, G. (2001). Six3 promotes the formation of ectopic optic vesicle-like structures in mouse embryos. *Dev Dyn* **221,** 342-9.

Laughon, A. (1991). DNA binding specificity of homeodomains. *Biochemistry* **30,** 11357-67.

Lee, H. H., and Frasch, M. (2005). Nuclear integration of positive Dpp signals, antagonistic Wg inputs and mesodermal competence factors during Drosophila visceral mesoderm induction. *Development* **132,** 1429-42.

Leppert, G. S., Yang, J. M., and Sundin, O. H. (1999). Sequence and location of SIX3, a homeobox gene expressed in the human eye. *Ophthalmic Genet* **20,** 7-21.

Li, X., Oghi, K. A., Zhang, J., Krones, A., Bush, K. T., Glass, C. K., Nigam, S. K., Aggarwal, A. K., Maas, R., Rose, D. W., and Rosenfeld, M. G. (2003). Eya protein phosphatase activity regulates Six1-Dach-Eya transcriptional effects in mammalian organogenesis. *Nature* **426,** 247-54.

Loosli, F., Koster, R. W., Carl, M., Krone, A., and Wittbrodt, J. (1998). Six3, a medaka homologue of the Drosophila homeobox gene sine oculis is expressed in the anterior embryonic shield and the developing eye. *Mech Dev* **74,** 159-64.

Loosli, F., Winkler, S., and Wittbrodt, J. (1999). Six3 overexpression initiates the formation of ectopic retina. *Genes Dev* **13,** 649-54.

Lopez-Rios, J., Gallardo, M. E., Rodriguez de Cordoba, S., and Bovolenta, P. (1999). Six9 (Optx2), a new member of the six gene family of transcription factors, is expressed at early stages of vertebrate ocular and pituitary development. *Mech Dev* **83,** 155-9.

Lopez-Rios, J., Tessmar, K., Loosli, F., Wittbrodt, J., and Bovolenta, P. (2003). Six3 and Six6 activity is modulated by members of the groucho family. *Development* **130,** 185-95.

Meyer, A., and Schartl, M. (1999). Gene and genome duplications in vertebrates: the one-to-four (-to-eight in fish) rule and the evolution of novel gene functions. *Curr Opin Cell Biol* **11,** 699-704.

Moore, L. A., Broihier, H. T., Van Doren, M., and Lehmann, R. (1998). Gonadal mesoderm and fat body initially follow a common developmental path in Drosophila. *Development* **125,** 837-44.

Murakami, Y., Ohto, H., Ikeda, U., Shimada, K., Momoi, T., and Kawakami, K. (1998). Promoter of mDMAHP/Six5: differential utilization of multiple transcription initiation sites and positive/negative regulatory elements. *Hum Mol Genet* **7,** 2103-12.

Niimi, T., Seimiya, M., Kloter, U., Flister, S., and Gehring, W. J. (1999). Direct regulatory interaction of the eyeless protein with an eye-specific enhancer in the sine oculis gene during eye induction in Drosophila. *Development* **126,** 2253-60.

Niiya, A., Ohto, H., Kawakami, K., and Araki, M. (1998). Localization of Six4/AREC3 in the developing mouse retina; implications in mammalian retinal development. *Exp Eye Res* **67,** 699-707.

Nishina, S., Kohsaka, S., Yamaguchi, Y., Handa, H., Kawakami, A., Fujisawa, H., and Azuma, N. (1999). PAX6 expression in the developing human eye. *Br J Ophthalmol* **83,** 723-7.

Ohto, H., Kamada, S., Tago, K., Tominaga, S. I., Ozaki, H., Sato, S., and Kawakami, K. (1999). Cooperation of six and eya in activation of their target genes through nuclear translocation of Eya. *Mol Cell Biol* **19,** 6815-24.

Ohto, H., Takizawa, T., Saito, T., Kobayashi, M., Ikeda, K., and Kawakami, K. (1998). Tissue and developmental distribution of Six family gene products. *Int J Dev Biol* **42,** 141-8.

Oliver, G., Mailhos, A., Wehr, R., Copeland, N. G., Jenkins, N. A., and Gruss, P. (1995). Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* **121,** 4045-55.

Oliver, G., Wehr, R., Jenkins, N. A., Copeland, N. G., Cheyette, B. N., Hartenstein, V., Zipursky, S. L., and Gruss, P. (1995). Homeobox genes and connective tissue patterning. *Development* **121,** 693-705.

Ostrin, E. J., Li, Y., Hoffman, K., Liu, J., Wang, K., Zhang, L., Mardon, G., and Chen, R. (2006). Genome-wide identification of direct targets of the Drosophila retinal determination protein Eyeless. *Genome Res* **16,** 466-76.

Ozaki, H., Nakamura, K., Funahashi, J., Ikeda, K., Yamada, G., Tokano, H., Okamura, H. O., Kitamura, K., Muto, S., Kotaki, H., Sudo, K., Horai, R., Iwakura, Y., and Kawakami, K. (2004). Six1 controls patterning of the mouse otic vesicle. *Development* **131,** 551-62.

Ozaki, H., Watanabe, Y., Takahashi, K., Kitamura, K., Tanaka, A., Urase, K., Momoi, T., Sudo, K., Sakagami, J., Asano, M., Iwakura, Y., and Kawakami, K. (2001). Six4, a putative myogenin gene regulator, is not essential for mouse embryonal development. *Mol Cell Biol* **21,** 3343-50.

Ozaki, H., Yamada, K., Kobayashi, M., Asakawa, S., Minoshima, S., Shimizu, N., Kajitani, M., and Kawakami, K. (1999). Structure and chromosome mapping of the human SIX4 and murine Six4 genes. *Cytogenet Cell Genet* **87,** 108-12.

Pasquier, L., Dubourg, C., Blayau, M., Lazaro, L., Le Marec, B., David, V., and Odent, S. (2000). A new mutation in the six-domain of SIX3 gene causes holoprosencephaly. *Eur J Hum Genet* **8,** 797-800.

Paululat, A., Breuer, S., and Renkawitz-Pohl, R. (1999). Determination and development of the larval muscle pattern in Drosophila melanogaster. *Cell Tissue Res* **296,** 151-60.

Pignoni, F., Hu, B., Zavitz, K. H., Xiao, J., Garrity, P. A., and Zipursky, S. L. (1997). The eye-specification proteins So and Eya form a complex and regulate multiple steps in Drosophila eye development. *Cell* **91,** 881-91.

Pineda, D., Gonzalez, J., Callaerts, P., Ikeo, K., Gehring, W. J., and Salo, E. (2000). Searching for the prototypic eye genetic network: Sine oculis is essential for eye regeneration in planarians. *Proc Natl Acad Sci U S A* **97,** 4525-9.

Purcell, P., Oliver, G., Mardon, G., Donner, A. L., and Maas, R. L. (2005). Pax6-dependence of Six3, Eya1 and Dach1 expression during lens and nasal placode induction. *Gene Expr Patterns* **6,** 110-8.

Ransom, R. (1979). The time of action of three mutations affecting Drosophila eye morphogenesis. *J Embryol Exp Morphol* **53,** 225-35.

Rayapureddi, J. P., Kattamuri, C., Steinmetz, B. D., Frankfort, B. J., Ostrin, E. J., Mardon, G., and Hegde, R. S. (2003). Eyes absent represents a class of protein tyrosine phosphatases. *Nature* **426,** 295-8.

Ribeiro, L. A., El-Jaick, K. B., Muenke, M., and Richieri-Costa, A. (2006). SIX3 mutations with holoprosencephaly. *Am J Med Genet A* **140,** 2577-83.

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P., Couronne, O., Hua, S., Smith, M. A., Zhang, P., Liu, J., Bussemaker, H. J., van Batenburg, M. F., Howells, S. L., Scherer, S. E., Sodergren, E., Matthews, B. B., Crosby, M. A., Schroeder, A. J., Ortiz-Barrientos, D., Rives, C. M., Metzker, M. L., Muzny, D. M., Scott, G., Steffen, D., Wheeler, D. A., Worley, K. C., Havlak, P., Durbin, K. J., Egan, A., Gill, R., Hume, J., Morgan, M. B., Miner, G., Hamilton, C., Huang, Y., Waldron, L., Verduzco, D., Clerc-Blankenburg, K. P., Dubchak, I., Noor, M. A., Anderson, W., White, K. P., Clark, A. G., Schaeffer, S. W., Gelbart, W., Weinstock, G. M., and Gibbs, R. A. (2005). Comparative genome sequencing of Drosophila pseudoobscura: chromosomal, gene, and cis-element evolution. *Genome Res* **15,** 1-18.

Roederer, K., Cozy, L., Anderson, J., and Kumar, J. P. (2005). Novel dominant-negative mutation within the six domain of the conserved eye specification gene sine oculis inhibits eye development in Drosophila. *Dev Dyn* **232,** 753-66.

Rubin, G. M., Yandell, M. D., Wortman, J. R., Gabor Miklos, G. L., Nelson, C. R., Hariharan, I. K., Fortini, M. E., Li, P. W., Apweiler, R., Fleischmann, W., Cherry, J. M., Henikoff, S., Skupski, M. P., Misra, S., Ashburner, M., Birney, E., Boguski, M. S., Brody, T., Brokstein, P., Celniker, S. E., Chervitz, S. A., Coates, D., Cravchik, A., Gabrielian, A., Galle, R. F., Gelbart, W. M., George, R. A., Goldstein, L. S., Gong, F., Guan, P., Harris, N. L., Hay, B. A., Hoskins, R. A., Li, J., Li, Z., Hynes, R. O., Jones, S. J., Kuehl, P. M., Lemaitre, B., Littleton, J. T., Morrison, D. K., Mungall, C., O'Farrell, P. H., Pickeral, O. K., Shue, C., Vosshall, L. B., Zhang, J., Zhao, Q., Zheng, X. H., and Lewis, S. (2000). Comparative genomics of the eukaryotes. *Science* **287,** 2204-15.

Ruiz Gomez, M., and Bate, M. (1997). Segregation of myogenic lineages in Drosophila requires numb. *Development* **124,** 4857-66.

Ruiz-Gomez, M. (1998). Muscle patterning and specification in Drosophila. *Int J Dev Biol* **42,** 283-90.

Ruiz-Gomez, M., Coutts, N., Price, A., Taylor, M. V., and Bate, M. (2000). Drosophila dumbfounded: a myoblast attractant essential for fusion. *Cell* **102,** 189-98.

Ruiz-Gomez, M., Romani, S., Hartmann, C., Jackle, H., and Bate, M. (1997). Specific muscle identities are regulated by Kruppel during Drosophila embryogenesis. *Development* **124,** 3407-14.

Rushton, E., Drysdale, R., Abmayr, S. M., Michelson, A. M., and Bate, M. (1995). Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. *Development* **121,** 1979-88.

Rushton, E., Drysdale, R., Abmayr, S. M., Michelson, A. M., and Bate, M. (1995). Mutations in a novel gene, myoblast city, provide evidence in support of the founder cell hypothesis for Drosophila muscle development. *Development* **121,** 1979-88.

Ruvkun, G., and Hobert, O. (1998). The taxonomy of developmental control in Caenorhabditis elegans. *Science* **282,** 2033-41.

Ryan, J. F., Burton, P. M., Mazza, M. E., Kwong, G. K., Mullikin, J. C., and Finnerty, J. R. (2006). The cnidarian-bilaterian ancestor possessed at least 56 homeoboxes: evidence from the starlet sea anemone, Nematostella vectensis. *Genome Biol* **7,** R64.

Sajithlal, G., Zou, D., Silvius, D., and Xu, P. X. (2005). Eya 1 acts as a critical regulator for specifying the metanephric mesenchyme. *Dev Biol* **284,** 323-36.

Sandmann, T., Girardot, C., Brehme, M., Tongprasit, W., Stolc, V., and Furlong, E. E. (2007). A core transcriptional network for early mesoderm development in Drosophila melanogaster. *Genes Dev* **21,** 436-49.

Sandmann, T., Jensen, L. J., Jakobsen, J. S., Karzynski, M. M., Eichenlaub, M. P., Bork, P., and Furlong, E. E. (2006). A temporal map of transcription factor activity: mef2 directly regulates target genes at all stages of muscle development. *Dev Cell* **10,** 797-807.

Sano, H., Renault, A. D., and Lehmann, R. (2005). Control of lateral migration and germ cell elimination by the Drosophila melanogaster lipid phosphate phosphatases Wunen and Wunen 2. *J Cell Biol* **171,** 675-83.

Santos, A. C., and Lehmann, R. (2004). Germ cell specification and migration in Drosophila and beyond. *Curr Biol* **14,** R578-89.

Sato, S., Nakamura, M., Cho, D. H., Tapscott, S. J., Ozaki, H., and Kawakami, K. (2002). Identification of transcriptional targets for Six5: implication for the pathogenesis of myotonic dystrophy type 1. *Hum Mol Genet* **11,** 1045-58.

Savard, J., Tautz, D., and Lercher, M. J. (2006). Genome-wide acceleration of protein evolution in flies (Diptera). *BMC Evol Biol* **6,** 7.

Savard, J., Tautz, D., Richards, S., Weinstock, G. M., Gibbs, R. A., Werren, J. H., Tettelin, H., and Lercher, M. J. (2006). Phylogenomic analysis reveals bees and wasps (Hymenoptera) at the base of the radiation of Holometabolous insects. *Genome Res* **16,** 1334-8.

Seimiya, M., and Gehring, W. J. (2000). The Drosophila homeobox gene optix is capable of inducing ectopic eyes by an eyeless-independent mechanism. *Development* **127,** 1879-86.

Seo, H. C., Curtiss, J., Mlodzik, M., and Fjose, A. (1999). Six class homeobox genes in drosophila belong to three distinct families and are involved in head development. *Mech Dev* **83,** 127-39.

Seo, H. C., Drivenes, Ellingsen, S., and Fjose, A. (1998). Expression of two zebrafish homologues of the murine Six3 gene demarcates the initial eye primordia. *Mech Dev* **73,** 45-57.

Seo, H. C., Drivenes, O., Ellingsen, S., and Fjose, A. (1998). Transient expression of a novel Six3-related zebrafish gene during gastrulation and eye formation. *Gene* **216,** 39-46.

Seo, H. C., Drivenes, O., and Fjose, A. (1998). A zebrafish Six4 homologue with early expression in head mesoderm. *Biochim Biophys Acta* **1442,** 427-31.

Serikaku, M. A., and O'Tousa, J. E. (1994). sine oculis is a homeobox gene required for Drosophila visual system development. *Genetics* **138,** 1137-50.

Shirokawa, J. M., and Courey, A. J. (1997). A direct contact between the dorsal rel homology domain and Twist may mediate transcriptional synergy. *Mol Cell Biol* **17,** 3345-55.

Silver, S. J., and Rebay, I. (2005). Signaling circuitries in development: insights from the retinal determination gene network. *Development* **132,** 3-13.

Spitz, F., Demignon, J., Porteu, A., Kahn, A., Concordet, J. P., Daegelen, D., and Maire, P. (1998). Expression of myogenin during embryogenesis is controlled by Six/sine oculis homeoproteins through a conserved MEF3 binding site. *Proc Natl Acad Sci U S A* **95,** 14220-5.

Spring, J., Yanze, N., Middel, A. M., Stierwald, M., Groger, H., and Schmid, V. (2000). The mesoderm specification factor twist in the life cycle of jellyfish. *Dev Biol* **228,** 363-75.

Staehling-Hampton, K., Hoffmann, F. M., Baylies, M. K., Rushton, E., and Bate, M. (1994). dpp induces mesodermal gene expression in Drosophila. *Nature* **372,** 783-6.

Stein, L. D., Bao, Z., Blasiar, D., Blumenthal, T., Brent, M. R., Chen, N., Chinwalla, A., Clarke, L., Clee, C., Coghlan, A., Coulson, A., D'Eustachio, P., Fitch, D. H., Fulton, L. A., Fulton, R. E., Griffiths-Jones, S., Harris, T. W., Hillier, L. W., Kamath, R., Kuwabara, P. E., Mardis, E. R., Marra, M. A., Miner, T. L., Minx, P., Mullikin, J. C., Plumb, R. W., Rogers, J., Schein, J. E., Sohrmann, M., Spieth, J., Stajich, J. E., Wei, C., Willey, D., Wilson, R. K., Durbin, R., and Waterston, R. H. (2003). The genome sequence of Caenorhabditis briggsae: a platform for comparative genomics. *PLoS Biol* **1,** E45.

Stierwald, M., Yanze, N., Bamert, R. P., Kammermeier, L., and Schmid, V. (2004). The Sine oculis/Six class family of homeobox genes in jellyfish with and without eyes: development and eye regeneration. *Dev Biol* **274,** 70-81.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) Software Version 4.0. *Mol Biol Evol* **24,** 1596-9.

Taylor, M. V. (2000). A novel Drosophila, mef2-regulated muscle gene isolated in a subtractive hybridization-based molecular screen using small amounts of zygotic mutant RNA. *Dev Biol* **220,** 37-52.

Tessmar, K., Loosli, F., and Wittbrodt, J. (2002). A screen for co-factors of Six3. *Mech Dev* **117,** 103-13.

Thisse, B., el Messal, M., and Perrin-Schmitt, F. (1987). The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. *Nucleic Acids Res* **15,** 3439-53.

Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22,** 4673-80.

Toy, J., and Sundin, O. H. (1999). Expression of the optx2 homeobox gene during mouse development. *Mech Dev* **83,** 183-6.

Toy, J., Yang, J. M., Leppert, G. S., and Sundin, O. H. (1998). The optx2 homeobox gene is expressed in early precursors of the eye and activates retina-specific genes. *Proc Natl Acad Sci U S A* **95,** 10643-8.

Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J., Gabor Miklos, G. L., Nelson, C., Broder, S., Clark, A. G., Nadeau, J., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W., Gong, F., Gu, Z., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z., Ketchum, K. A., Lai,

Z., Lei, Y., Li, Z., Li, J., Liang, Y., Lin, X., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B., Sun, J., Wang, Z., Wang, A., Wang, X., Wang, J., Wei, M., Wides, R., Xiao, C., Yan, C., et al. (2001). The sequence of the human genome. *Science* **291,** 1304-51.

Wallis, D. E., Roessler, E., Hehr, U., Nanni, L., Wiltshire, T., Richieri-Costa, A., Gillessen-Kaesbach, G., Zackai, E. H., Rommens, J., and Muenke, M. (1999). Mutations in the homeodomain of the human SIX3 gene cause holoprosencephaly. *Nat Genet* **22,** 196-8.

Wargelius, A., Seo, H. C., Austbo, L., and Fjose, A. (2003). Retinal expression of zebrafish six3.1 and its regulation by Pax6. *Biochem Biophys Res Commun* **309,** 475-81.

Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., Agarwala, R., Ainscough, R., Alexandersson, M., An, P., Antonarakis, S. E., Attwood, J., Baertsch, R., Bailey, J., Barlow, K., Beck, S., Berry, E., Birren, B., Bloom, T., Bork, P., Botcherby, M., Bray, N., Brent, M. R., Brown, D. G., Brown, S. D., Bult, C., Burton, J., Butler, J., Campbell, R. D., Carninci, P., Cawley, S., Chiaromonte, F., Chinwalla, A. T., Church, D. M., Clamp, M., Clee, C., Collins, F. S., Cook, L. L., Copley, R. R., Coulson, A., Couronne, O., Cuff, J., Curwen, V., Cutts, T., Daly, M., David, R., Davies, J., Delehaunty, K. D., Deri, J., Dermitzakis, E. T., Dewey, C., Dickens, N. J., Diekhans, M., Dodge, S., Dubchak, I., Dunn, D. M., Eddy, S. R., Elnitski, L., Emes, R. D., Eswara, P., Eyras, E., Felsenfeld, A., Fewell, G. A., Flicek, P., Foley, K., Frankel, W. N., Fulton, L. A., Fulton, R. S., Furey, T. S., Gage, D., Gibbs, R. A., Glusman, G., Gnerre, S., Goldman, N., Goodstadt, L., Grafham, D., Graves, T. A., Green, E. D., Gregory, S., Guigo, R., Guyer, M., Hardison, R. C., Haussler, D., Hayashizaki, Y., Hillier, L. W., Hinrichs, A., Hlavina, W., Holzer, T., Hsu, F., Hua, A., Hubbard, T., Hunt, A., Jackson, I., Jaffe, D. B., Johnson, L. S., Jones, M., Jones, T. A., Joy, A., Kamal, M., Karlsson, E. K., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* **420,** 520-62.

Wawersik, M., Milutinovich, A., Casper, A. L., Matunis, E., Williams, B., and Van Doren, M. (2005). Somatic control of germline sexual development is mediated by the JAK/STAT pathway. *Nature* **436,** 563-7.

Wawersik, S., and Maas, R. L. (2000). Vertebrate eye development as modeled in Drosophila. *Hum Mol Genet* **9,** 917-25.

Weasner, B., Salzer, C., and Kumar, J. P. (2007). Sine oculis, a member of the SIX family of transcription factors, directs eye formation. *Dev Biol* **303,** 756-71.

Weigmann, K., Klapper, R., Strasser, T., Rickert, C., Technau, G., Jackle, H., Janning, W., and Klambt, C. (2003). FlyMove--a new way to look at development of Drosophila. *Trends Genet* **19,** 310-1.

Wolpert, L. (1998). "Principles of development." Current Biology ;
Oxford University Press, London
Oxford.

Xu, P. X., Adams, J., Peters, H., Brown, M. C., Heaney, S., and Maas, R. (1999). Eya1-deficient mice lack ears and kidneys and show abnormal apoptosis of organ primordia. *Nat Genet* **23,** 113-7.

Xu, P. X., Zheng, W., Huang, L., Maire, P., Laclef, C., and Silvius, D. (2003). Six1 is required for the early organogenesis of mammalian kidney. *Development* **130,** 3085-94.

Yanagi, K., Mizuno, T., Tsuyama, T., Tada, S., Iida, Y., Sugimoto, A., Eki, T., Enomoto, T., and Hanaoka, F. (2005). Caenorhabditis elegans geminin homologue participates in cell cycle regulation and germ line development. *J Biol Chem* **280,** 19689-94.

Yanowitz, J. L., Shakir, M. A., Hedgecock, E., Hutter, H., Fire, A. Z., and Lundquist, E. A. (2004). UNC-39, the C. elegans homolog of the human myotonic dystrophy-associated homeodomain protein Six5, regulates cell motility and differentiation. *Dev Biol* **272,** 389-402.

Zhou, X., Hollemann, T., Pieler, T., and Gruss, P. (2000). Cloning and expression of xSix3, the Xenopus homologue of murine Six3. *Mech Dev* **91,** 327-30.

Zhu, C. C., Dyer, M. A., Uchikawa, M., Kondoh, H., Lagutin, O. V., and Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. *Development* **129,** 2835-49.

Zilversmit, M., O'Grady, P., and Desalle, R. (2002). Shallow genomics, phylogenetics, and evolution in the family Drosophilidae. *Pac Symp Biocomput,* 512-23.

Zou, D., Silvius, D., Fritzsch, B., and Xu, P. X. (2004). Eya1 and Six1 are essential for early steps of sensory neurogenesis in mammalian cranial placodes. *Development* **131,** 5561-72.